

APPLIED STATISTICS

A Journal of the
Royal Statistical Society

EDITORS

LEONARD H. C. TIPPETT

AND

DONALD G. BEECH



VOLUME FOUR

Oliver and Boyd Ltd.
London: 39A Welbeck Street, W.1
Edinburgh: Tweeddale Court

1955

CONTENTS

No. 1. MARCH	<i>Page</i>
The Productivity Index in Active Service <i>Stafford Beer</i>	1
A Method of Graphing Sales Figures for Executive Use <i>Albert W. Swan</i>	15
An Application of Linear Programming to the Selection of Raw Materials <i>Noel Williams</i>	22
The Comparison of Means of Industrial Time Series <i>Geoffrey H. Jowett</i>	32
Some Makeshift Methods of Analysis Applied to Complex Experimental Results <i>G. P. Davies and G. W. Sears</i>	47
Letter to the Editor: A Point in the Construction of Price Index Numbers	64
Question and Answer	65
Industrial Applications Section of the Royal Statistical Society: List and Reports of Meetings	67
Study Section of the Royal Statistical Society: Reports of Meetings ..	70
Book Reviews:	
<i>The Design and Analysis of Industrial Experiments</i> , O. L. Davies (Editor)	71
<i>Biometrika Tables for Statisticians</i> , E. S. Pearson and H. O. Hartley (Editors)	72

No. 2. JUNE

Reservoirs of Infection of Pulmonary Tuberculosis in Warwickshire <i>Marjorie Jessop</i>	73
Multiple Regression between Variables measured at the Points of a Plane Sheet <i>Geoffrey H. Jowett</i>	80
Calculation by Accumulation .. <i>P. A. Daniell and F. D. K. Liddell</i>	90
A Simple Non-Parametric Graphical Test for the Significance of a Difference <i>Wladyslaw Z. Billewicz</i>	97
Multivariate Analysis <i>John Wishart</i>	103
Letter to the Editor: Theoretical and Practical Workers in Statistics ..	117
Notes and Comments	120

CONTENTS (continued)

	Page
Industrial Applications Section of the Royal Statistical Society: List and Reports of Meetings	122
Study Section of the Royal Statistical Society: Reports of Meetings ..	126
Book Reviews:	
<i>Elements of Statistics</i> , H. C. Fryer	127
<i>Handbook of Graphic Presentation</i> , Calvin F. Schmid	127
<i>The Design of Social Research</i> , R. L. Ackoff	128
<i>Statistical Analysis in Chemistry and the Chemical Industry</i> , Carl A. Bennett and Norman L. Franklin	129
<i>Quality Control through Statistical Methods</i> , Norbert Lloyd Enrick ..	130
<i>Productivity, a Critique of Current Usage</i> , Lewis A. Maverick ..	130
<i>A Concise History of Mathematics</i> , D. J. Struik	131
Publications Received	132

No. 3. NOVEMBER

Current Trends in National Income Statistics	<i>Christopher T. Saunders</i>	133
Fundamental Problems in the Subjective Appraisal of Foodstuffs	<i>Roland Harper</i>	145
Sales Forecasting: An Evaluation of Some Practical Results	<i>David J. Liston, MBE</i>	162
Level of Quality for Minimum Cost of Manufacture to a Specification	<i>John H. Hepburn</i>	173
Measuring the Readership of Newspapers and Magazines	<i>Eric Shankleman</i>	183
The Determination of Branch Staff Establishments in a Multiple Shop Organisation	<i>Bernard T. Ramm</i>	195
A Design for Testing Several Treatments under Controlled Environmental Conditions	<i>Geoffrey M. Clarke</i>	199
Notes and Comments		207
Industrial Applications Section of the Royal Statistical Society: List and Reports of Meetings		208
Study Section of the Royal Statistical Society: List and Reports of Meetings		211
Book Review:		
<i>Biomathematics: The Principles of Mathematics for Students of Biological Science</i> , Cedric A. B. Smith		215
Publications Received		219

THE PRODUCTIVITY INDEX IN ACTIVE SERVICE*

STAFFORD BEER

Samuel Fox and Co. Ltd.

Mr Beer describes a productivity index that has been used in a steel works in production planning and control. He deals with the technical and administrative as well as with the statistical aspects of his subject and discusses additional management uses of the index. The subject is treated from the point of view of a works with a large number of varied and changing products.

What is the Productivity Index?

There are many ways of measuring productivity, some more effective than others. 'Output per man-hour', for instance, is not a very satisfactory measure because it takes no account of input factors other than time and labour. A rather different approach, recommended by the Anglo-American Council on Productivity,¹ compares the *objective* time for a process (the time it really should take when everyone is working hard and the process has been brought to its ultimate level of technical efficiency) with the *actual* time for that process (the time it really did take). Thus:

Productivity Index (PI) = $100 \times \text{objective time} \div \text{actual time}$.

The advantage of this form of index lies in its comprehensiveness; it takes everything into account, and provides a common measure of productivity for any mixture of machines, processes, and products.

The main difficulty in producing such an index lies in obtaining a good estimate for the numerator. The best basis is a mathematical formula of the process concerned. If such a formula can be constructed it is a simple matter to apply any set of variables to it, and to read off the objective time for the job to which the variables relate.

Here is a simple formula for a cold rolling mill in the steel industry. We have a number c of coils of steel strip, b inches in breadth, to be rolled to g inches in thickness, and weighing w pounds in all. The steel has a density of d pounds per cubic inch, and the average speed of the mill is v feet a minute. It takes H minutes handling time to mount and take off each coil; and h minutes handling time is further consumed each time the strip is passed through the mill, which happens p times. Then the objective time will be given by:

$$t = p \left(\frac{w}{12dbgv} \right) + chp + cH$$

* This article is based on a paper given at a conference of the Industrial Applications section of the Royal Statistical Society held at Cranfield in September 1953.

This formula will be correct whatever the actual numbers, and so we can feed into it the numerical details for any job. It simply calculates the theoretical length of the strip, which is then divided by the theoretical speed to give a total 'run through' time. This has to be multiplied by p , which may be either an empirical estimate or a theoretical calculation. The allowances h and H are best obtained by time study. Dividing 100 times the result by the time the job actually took will give a percentage measure of productivity. This is the index for that job.

Analysing Output in Terms of the Index

If we now take a record of production and calculate the productivity index of every job contained in that record, a set of results will be obtained to give an extremely good picture of what is happening on the shop floor. Remember that these results can be compared with each other, whatever the product of process may be, because objective times have all been calculated in relation to an ultimate standard of attainment.

There are, however, some dangerous pitfalls in choosing a statistical sample of production for this purpose. In collecting data from a whole department, the temptation to tabulate them as a matter of course in groups with which everyone is already familiar should be resisted. Products which are associated in everyone's minds for marketing reasons can not necessarily be grouped together as having commensurable levels of productivity. Equally, machines with similar levels of *production* may operate at quite different productivities. For this sort of reason there may be no genuine basis for regarding a specially chosen sample of work as typical of production: it may merely be typical of the chooser's prejudice. Random sampling is also unwise at this stage because it would have to assume a framework within which to operate. For example, since the precise weighting over an unknown period or unrecognised cycle of any one unknown factor may be of crucial importance, random sampling may not yield sufficient information. Unfortunately, the use of balanced designs is nearly always impracticable in plants working under the pressure of full production; there may be unacceptable delays in obtaining the necessary combinations of factors, and replication may be literally impossible. In short, any sort of experimentation is difficult; we must use data relating to the situation as it already happens to be.

The only practicable procedure in making the primary analysis seems to be to lift the whole body of production records on to statistical data sheets, covering as long a period as seems needed to take into account any cyclical variation which may exist. Fixing this period means making an assumption, and it is well to test the proposal beforehand. In one department a six-week period proved adequate; in another, it was more than six months.

When a mass of data has been collected in this way, and operating schedules have been built up to supply the values of variables required

to work out the formula for objective time, it can be collated ready for the calculation of indices. This is accomplished on desk-calculating machines which work out the formula according to a set routine which the operator quickly learns. Thousands of commensurable indices then become available for examination. They probably range from zero (a measure of complete breakdown) to over one hundred (a measure of the exceeding of allegedly maximum performance). In a big department these data will need to be handled on punched cards; even so, a full-scale multifactor analysis of variance would be a virtually impracticable task.

The means adopted to introduce order into this chaos vary and are difficult to describe in detail. Firstly, however, we break up the data into 'atomic' sub-groups. This means separating products, processes, and machines, and gradually increasing the number of divisions until the whole situation has been factorised. The stage of atomic factorisation is recognised both by the logic of the situation, and by the statistical factor of productive homogeneity within each group as determined by significance testing. That is, each little group of indices has a homogeneous frequency distribution with a single mode, which represents a distinct 'parcel' of output.

The second stage is to synthesise the 'atoms' into 'molecules'. By this I mean that similar atomic sub-groups may be coalesced to form 'molecular' groups, use being made of the orthodox techniques of significance testing. When a molecular group is completed (that is, when no further atomic parts can be merged into it), the effective cause of its significant difference from other groups can be determined as the common factor of its constituent atoms. This analogy from chemistry is useful, because it indicates the sort of grouping that is valid. Only those atomic groups which there is logical reason to suppose really hang together should be tested against each other.

To elaborate this point a little, let us consider what might happen in practice. Product A and product B are manufactured on the same machine by the same operators. A and B themselves are almost the same: they differ only, say, in that B is half an inch wider than A. These are atomic groups already isolated by the primary analysis, and it is reasonable to ask whether these atoms belong to the same molecule. Unless some critical limit exists in that half-inch of width, there seems no good reason why they should not be classified together. We therefore compare the parameters of each atom, A and B, by *t*- and *F*-tests and find no significant difference between them. The molecular group (A + B) is then established. If, on the other hand, there were a significant difference, we should be driven to the conclusion that for some reason—possibly the existence of a critical width, possibly of another and unrecognised factor—the two atomic groups must be kept separate as belonging to different molecules. The attempt to merge two atoms into a single molecule may even be justified when A and B refer to different machines if there is some strong link between them. For

instance, the two machines may both operate within the same scheme of wage incentives. If they have been rated for this purpose in a way calculated to tie comparable rewards to comparable efforts, it will not be surprising if the productivity distributions are also tied together.

We must not, however, be misled by a coincidental similarity in statistical parameters to merge atoms which clearly cannot belong to the same molecule. Consider two products X and Y, which have nearly the same mean productivity and variance. Product X is a coil of wire resulting from a drawing process; product Y is a small forging produced in a press. Significance tests will indicate that X and Y belong to the same population, whereas they quite obviously do not. This pitfall is particularly important in this type of work but it is not unfamiliar in statistics generally. In fact it is a sort of 'nonsense correlation' and merely indicates that methods of statistical inference are not independent of other forms of logic. The building of a realistic statistical framework on which data can be hung always calls for insight and diligence.

Summing up, we may say that the output of a whole department can be subjected to this analysis. As a result, that output will be split into a number of 'productivity groups', each of which will be defined in two ways. Firstly, it refers to certain products undergoing certain processes on certain machinery. Secondly, its most important statistical parameters are known: there is a mean productivity with a known variance. Our insight into the department is almost complete, and is expressed in a statistical form convenient to handle.

A Necessary Transformation

Before we can follow the development of this productivity system into any sort of action, a statistical problem must be faced. The statistical parameters mentioned are not sufficient unless the distribution of productivities within each group is also stated or is normal. If, as has been so far assumed, the analysis is carried out on the raw indices themselves there is a strong tendency for a distribution skewed to the left. The reason for this is easy to see.

Suppose that the mean PI is 90 with a standard deviation of 10. If, for the moment, we assume the distribution to be normal, 31.73 per cent of the total area will fall outside the one-sigma limits of 90 ± 10 . Hence nearly 16 per cent of occurrences will fall beyond the 100 value. But the index is so constructed that a PI of 100 cannot be exceeded. The result is that productivity rising above the mean of 90 will tend asymptotically to the 100 maximum and the 16 per cent of occurrences which would in normal circumstances spread beyond the 100 mark will agglomerate between the mean and 100, together with the 34 per cent of occurrences contained in that 90-100 area as of right. This 'agglomeration effect' is inherent in the nature of the index; if it is not present, it will tend to become present as efficiency increases. Even when the mean PI is so low that the spread is in no danger of

agglomerating against the 100 limit, the resistance of that ultimate inhibition may be felt and may produce an asymmetrical distribution. I have found these productivity distributions particularly susceptible to this form of skewness, which can easily give rise to a 'precipice' effect.

The PI distribution is similar in appearance to the binomial distribution, and the transformation that would be used for the latter, that is the inverse sine transformation, eliminates the objectionable agglomeration effect, as may be seen in Fig. 1, and makes the distribution substantially normal. Accordingly, every productivity index is transformed to the angle θ , according to the expression:

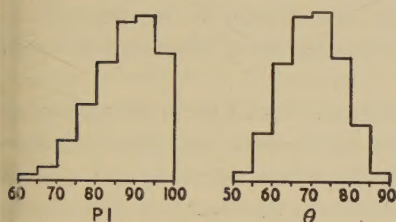


FIG. 1. Distributions illustrating how the 'agglomeration effect' of indices against an ultimate limit is dispersed by the inverse sine transformation which makes the distribution approximately normal.

$$\theta = \sin^{-1} \sqrt{(x/X)}$$

where x is the PI and X is the inhibiting limit, which in this case is 100. Tables are available² from which θ may be read directly for the ratio x/X ($= x/100$).

The mathematical account of this³ and of another, similar,⁴ transformation is interesting; but the practical conclusion is that all indices should be transformed from the start, when they will become more amenable to treatment as members of normal distributions.

How to Keep the Groups Running

The picture so far is a static one, correct at the time of primary analysis. But what of the future? If the analysis has been properly done, the analyst expects that the productivity groups will not promptly disintegrate. But productivity does change; so it is likely that the groups will change in mean value as time goes by. Hence, to make the picture dynamic it will be necessary to *control* the parameters of the groups as a matter of routine.

This may be done by means of control charts, one of which is drawn up for each productivity group. The data are obtained from a sampling programme according to which every shift in a four-weekly period is numbered. At the end of each period, a group of these code numbers is randomly selected and sent to the recording section of the production department, where the output details of the shifts thereby identified are entered on a form. The data constitute a random sample of each group; but the adequacy of the sample sizes is difficult to forecast because the amount of production time spent under each group heading varies from period to period. In order to ensure that the sample size cannot possibly fall below five per cent of the output, it has been necessary to fix the number of shifts drawn in each group independently and at a level which in practice results in an average sample size for the whole department of twelve per cent.

The data sheets on which the periodic samples are compiled are then subjected to the computation schedule mentioned earlier. This calculates the index for each shift, and its 'inverse sine', and finally produces the sample means and ranges in terms of θ . The charts on which these figures are plotted are based on parameters of the original but transformed indexical groups, with inner limits at $1.96\sigma/\sqrt{n}$ and outer limits at $3.09\sigma/\sqrt{n}$. Thus they are orthodox charts from the point of view of theory, although a novel graphical method of plotting them has been devised. This technique⁵ enables the charts to be massed in batteries of uniform ranks, thereby facilitating their handling in large numbers.

By operating these statistical controls in the usual way, we can keep the picture of productivity permanently up-to-date, altering the controlled parameters (after special analysis) whenever they go out of control. In practice, of course, the statistician in charge of the charts will normally notice a tendency towards significant change. Except in unusual cases, which are probably heralded, such changes do not happen overnight. When the periodic sample regularly gives a value outside the inner limits, the statistician will take some measure to reach an early decision; he will at least ask for intermediate samples so that the incipient change may be watched.

As far as trends are concerned, various techniques of recognition are particularly appropriate and deserve mention. Firstly, there is the chi-squared testing of the frequencies attached to plus and minus values of samples about the control mean. Secondly, there is the ratio test developed by von Neumann which recognises sequences, a form of which (complete with significance tables) has been evolved by Baines.⁶ The application of this is well treated in a useful American publication.⁷ Again, there is the expedient of hastening decisions on control charts by contracting the outer limits. This may be done, without waste, by increasing the sample to what is by definition⁸ its most economical size.

So far, we have been considering the control of established groups; this leaves the question of new groups, which may be constructed from data relating to new machines or new processes. Assuming that a new group has been identified, and validated in the way earlier described, it can either be regarded as a new discrete entity, or (if the proper conditions are fulfilled) it can be coalesced with an existing group. In an established and comprehensive system, it is likely that the new group will correspond markedly with one or two existing groups.

The decision as to whether two groups can be coalesced will be based as before on the logic of the situation; statistically, it will require the acceptance or rejection of the null hypothesis as to the identity of their two means. Further, it is desirable that a significance test aimed at making this decision should pronounce on the question with all speed, and this suggests the use of a sequential test. Such a test,⁹ derived from Wald's theory of sequential analysis, has been used. It does, however, present difficulties in this particular context and a simple sequential

version of the u -test may be preferable. This tests undistributed samples into an existing group variance according to the statistic $u = (\bar{x} - \mu) \sqrt{n}/\sigma$. All that is required is to accumulate the results of the new product as time goes by, plotting Σx against n for a chosen significance of u . This procedure is extremely economical in sample sizes.

A Valuable Modification

The general Productivity Index so far described is a ratio between the *objective* and the *actual* time. Its movements indicate changes, the nature of which it is impossible to assess without further information. A refinement may be introduced by creating a third time concept called the *standard* time. This will normally fall between the other two times. It is arranged to modify the full objective time, which represents all factors as operating at maximum efficiency, by relaxing the stringency of each of those factors. The criterion for this procedure is the extent to which it is technically difficult to achieve the objective levels. For example, the speed of the process may have to be scaled down to accord with current limitations in lubrication or temperature control; ratings for handling may be lowered to allow for imperfections in the handling techniques; and, most of all, an allowance will have to be added to allow for the current level of technical stoppage to the plant.

This standard time, like the objective time, is worked out from a modified form of the original formula. The point is that the difference between objective and standard is a measure of current technical shortcomings. The ratio of the two times provides a new index thus:

Current Technical Index (CTI)

$$= 100 \times \text{objective time} \div \text{standard time.}$$

Movements in this index provide a measure of change in *technical* efficiency.

Since, moreover, the technical influences on the general Productivity Index have been isolated in this way, it follows that the element of change still unaccounted for must be due to non-technical operating factors. This element is the difference between the standard and actual times and is a measure of labour utilisation. This yields the index:

Labour Utilisation Index (LUI)

$$= 100 \times \text{standard time} \div \text{actual time.}$$

Movements in this index provide a measure of change in *labour* efficiency.

If these two subsidiary indices are multiplied together and divided by 100 the general Productivity Index is the result. Hence, by keeping records of these indices as well, we provide ourselves with an insight into the nature of productivity changes as well as a measure of their movements. This procedure is recommended by the originators of this productivity measure.¹

Forecasting Production Times

Once the whole output of a production department has been analysed in terms of the triple index of productivity, organised into statistical groups, and subjected to routine statistical control, real action can begin. The fundamental notion behind all the uses to which this set-up has been put is the notion of a production forecast. However complicated the job, and even if it is a hitherto untried manufacture, this forecast is now quite easy to make.

The specification of the job, together with the appropriate scheduled factors, are fed into the mathematical formulae for the objective time. The objective time is a useful limiting concept in that we have a target and a measure of ultimate capacity, but is unrealistic as a forecast. Suppose that, postulating ultimate labour efficiency of one hundred per cent, it is desired to know the standard time; then it is only necessary to multiply the objective time by the reciprocal of the Current Technical Index. To find the realistic production forecast, duly weighted for both technical and labour factors, we multiply the objective time by the reciprocal of the Productivity Index itself. The PI figure used is the mean value of the productivity group to which the job belongs. Its current validity is guaranteed by the statistical control.

In what way is this technique of forecasting preferable to the more primitive method of taking actual average times and making corrections to suit changed circumstances? Firstly, in jobbing production it might be necessary to hold a library of perhaps 4 000 such actual averages, which can now be replaced by about 70 PI control groups. Secondly, the keeping of control charts for the separate jobs would be a superhuman task; experience shows that these historical standards rapidly become out of date and cannot be kept up-to-date by a staff of reasonable size. Thirdly, a considerable proportion of jobs may be new work, previously unattempted. Someone then has to guess which standard to use. By the new method, a new job can be fitted into its proper PI group on sight, because the catalogue of groups collectively exhausts the possibilities of production. An experiment carried out on the ability of historical standards to forecast the actual times of jobbing work on samples of five shifts gave an error of seventy per cent. This was partly due to the standards being out of date, and partly due to wrong judgements as to which would be the nearest standard to use. The error incurred on similar samples when the new technique was used was less than three per cent.

The accuracy of these forecasts is extremely high and it is worth considering the main reason why this should be so, for, one might argue, the whole calculation is based on a rather arbitrary estimate of objective time. Suppose, for instance, the handling allowances made at that stage were quite unrealistic? The answer is that this simply does not matter. If the primary analysis has incorporated an error of this sort in the objective, the index will also incorporate the error. When the forecast is made, the same mistake is made in the objective, and is

cancelled out when the index is applied as a weight. To demonstrate:

$$\begin{aligned}\text{Forecast time} &= \text{objective} \times 1/\text{PI} \\ &= \text{objective} \times \text{actual/objective} \\ &= \text{actual}.\end{aligned}$$

Any constant added to or subtracted from the objective in error or by misjudgement will not affect this forecast.

The formulae in Fig. 2 gather together all the threads of this discussion to this point, and consider them from the standpoint of their main function—forecasting. This is an actual example, which extends to its full state the system as applied to steel strip manufacture. The model is applicable to all continuous processes and not merely to rolling as described in the introductory section.

$$T = 100t_o/P$$

Where
$$P = 100 \frac{t_o}{t_s} \cdot \frac{t_s}{t_a} \quad (t_o \leq t_s \leq t_a)$$

$$t_o = p \left(\frac{w}{12dbgv} \right) + chp + cH$$

$$t_s = p \left(\frac{w}{12dbgv'} \right) + ch'p + cH' + zc + Zc$$

and
$$t_a = re/i$$

θ is normally distributed for:

$$\theta = \sin^{-1} \sqrt{(P/100)}$$

KEY:

T = forecast time, P = productivity index.

Inferred Quantities

t_o = objective time

t_s = standard time

t_a = actual time

Given Quantities

i = time involved

e = time expected to work

r = time recorded as worked

Variables

p = number of passes

w = weight in lb.

b = breadth of strip

g = thickness of strip

c = number of coils

Constants for any given job

d = density in lb. per cu. in.

v = objective } process speed

v' = standard }

h = objective } handling time

h' = standard } per pass

H = objective } handling time

H' = standard } per coil

z = regular } technical stops

Z = irregular } estimated per coil

Note. Slight adjustments are also made in practice to compensate for time losses due to acceleration and deceleration.

Fig. 2. A display of the quantities and formulae used in developing the productivity index and forecast of the time to pass a batch of coils of steel strip through a given process on a given machine.

Production Control

I have shown how heterogeneous output may be measured in terms of a triple index of productivity, how the work of a factory may be

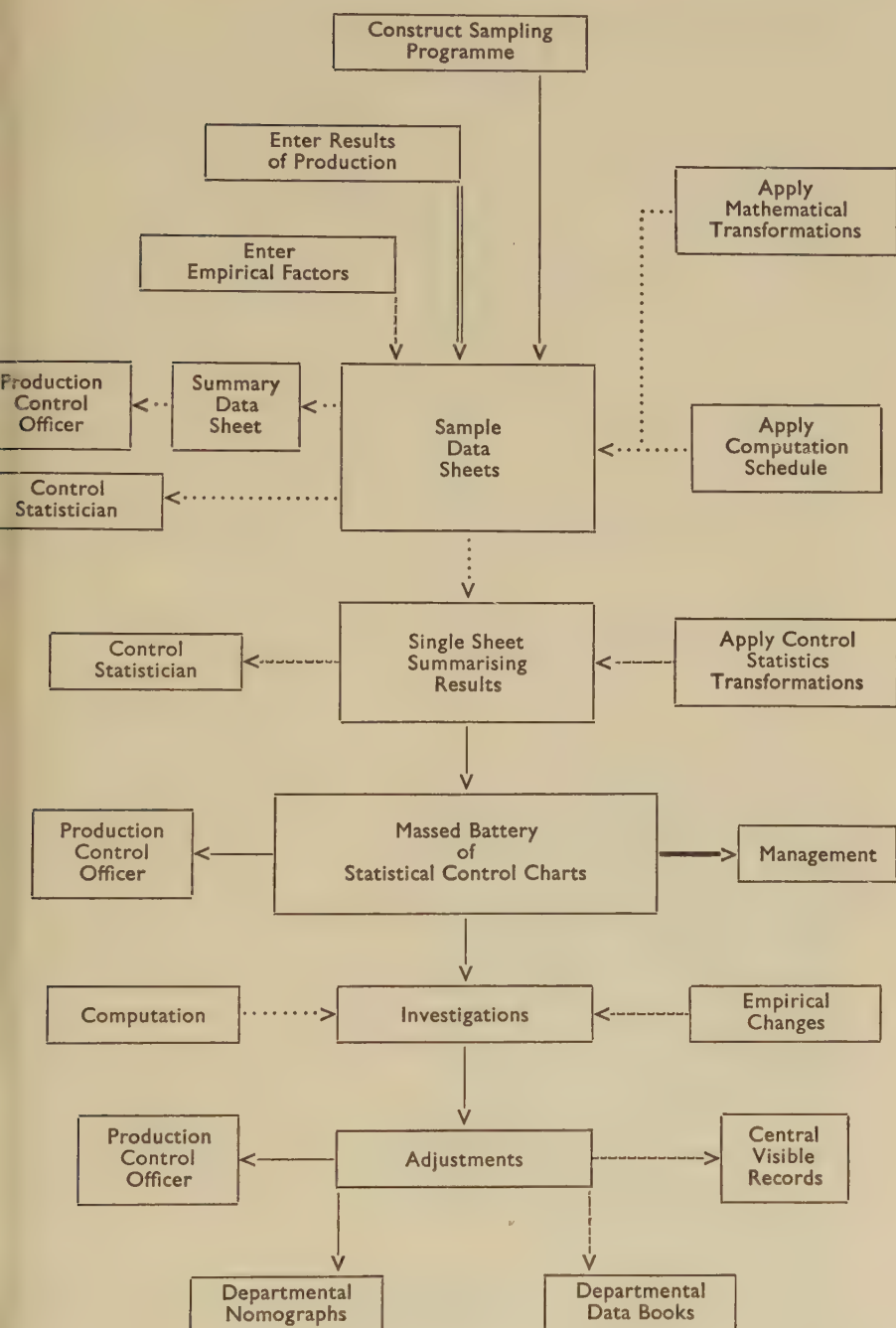
classified in terms of this measure, and how that measure may be used to convert theoretical expectations of production into dependable long-term forecasts. I have also described a means of controlling this system, once installed, with the object of ensuring that the forecasts made are consistent with latest developments on the shop floor. There are refinements, not mentioned, but I have described the general structure of a scheme which has been found to be practicable. A tool of some power has been forged, which, for the reasons given two paragraphs ago, is independent of the uncertainties of time study, and which has many applications remote from its immediate provision of forecasts.

It was, indeed, to facilitate production control that the scheme was actually devised. Thorough control of planning requires a thorough understanding of production. It is this which is so hard to obtain in a diversified works, and this which once obtained is so hard to formulate exactly. The technique achieves both these objects. It remains to settle the mechanics of programming on this basis, which resolves into a question of how to compute these scientific forecasts in an ordinary clerical office.

That problem of the day-to-day calculation of forecasts was overcome by a multi-stage nomograph. This was designed and built specifically to calculate, from an order for strip, a forecast of production time required. The particular chart is six feet long and consists of eleven interlocking nomographic cycles, but other charts have been produced to solve different formulae. All are covered with sixteenth-inch transparent plastic, and transparent rules with sharply pointed ends are used to move across them; this, it is found, gives an error of the order of only one per cent. Constants are built-in, and the operator has merely to impart the information on the order form, and the value of the variables contained in his schedules. The nomograph then solves the equation for objective time, and adjusts it by the appropriate productivity constants. Thus the objective, the standard, and the actual times may all be forecast for a given job in something under a minute.

These three times are the raw material for whatever answers may be required from a production control scheme. Budgets and forecasts based on a variety of assumptions can be composed from them and the differences between them can be interpreted in research, labour, financial, or commercial contexts as required.

The production control scheme within a works department is thus able to operate as an independent entity, except for the functional control of its statistical mechanism. This requires simply that significant changes detected on the centralised control charts are immediately 'fed back' to the nomographic scales dealing with productivity weights. By this means, departmental forecasts are always realistic, being based on the very latest evidence from the shop floor. A diagrammatic representation of the central process by which a departmental scheme is functionally controlled is provided in Fig. 3.



RESPONSIBILITIES



FIG. 3. A departmental scheme of production control.

Detailed Production Studies

It sometimes happens that more detailed insight into production is required than the triple index provides. For instance, it may be necessary to isolate certain factors in order to examine alleged improvements in production technique, or to see exactly where losses in productivity are occurring. In both types of enquiry, those immediately concerned are tendentially inclined to use one of the spurious measures (say cost or weight) readily to hand, for one can usually be found to lend verisimilitude to any case. The index, however, is a most useful comprehensive measure from which to start, and by various devices we have been able to elaborate the information contained in the triple index to the point where an answer is forthcoming to both these questions.

When the productivity is already known, and the losses are moreover already allocated (as they are) between technical and labour factors it merely remains to fill in this framework. The problems of method involved are then more readily solved than by traditional work study investigation. Statistical methods are admirable in this context; those suggested by Abruzzi,⁷ and the Tippett 'snap-reading' technique¹⁰ have proved especially useful.

Routine Reports to Management

The form in which statistical conclusions can best be presented to management has properly received much attention. For a scheme of this sort, the reporting is done in two parts. First, an indexed book is provided as a permanent 'dictionary of productivity', which is kept up to date by amendments issued whenever a controlled mean is adjusted. Reference to this book provides management with the history of productivity for any job together with its present level. Secondly, a report is issued each time a routine sample is drawn, on which are recorded those groups which have departed from the controlled mean to the extent of $1.96\sigma/\sqrt{n}$ or more. The probability level associated with such a sample average is designated on a 'scale of importance' which reaches 10 at the 1-in-40 level and somewhat overruns the 1-in-100 level at a reading of 16. In addition to this scale, a second or 'variability scale' is quoted which derives⁵ from the sample ranges. Each of these scales is calibrated to interpret significance levels to management in ordinary English, using language calculated to inspire action.

Cost Accounting

Without entering into technical details¹¹ which would not be germane to this article, it will be clear that actual times inferred from this forecasting system can be used in the establishment of accounting standards. The experiment has, in fact, been tried, when the processes described in this paper alone were used to generate (on desk calculators) 120 000 machine/job standards. In the outcome, the discrepancy between the standard cost thus computed and the actual cost of production

in the very first month's application was less than one per cent.

Apart from its application to standard costing, the productivity index system can offer assistance in analysing the cost of flexibility: a desirable factory 'commodity' which has nevertheless to be bought. The price of flexibility is an 'opportunity cost' of a profound sort, and one which present-day cost accounting practice is probably not in a position to assess at all accurately. I say this because to cost an opportunity foregone requires the costing of an ideal situation; but the ideal situation cannot be identified until *after* it has been costed. This is because the raw measures of output used by accountants are themselves immiscible, and are often put into arbitrary groups.

Triple indices of productivity, however, that are properly understood, systematically ordered, and statistically controlled, can be used very easily to construct an estimate of this opportunity cost in terms of output capacity. They can also apportion this abandoned capacity among the various reasons for which flexibility is demanded; for example, in achieving especially high quality, in implementing a liberal personnel policy, or in meeting special and sudden demands from important customers. Such decisions are matters of company policy. But a statistical control scheme is surely uniquely situated to provide evidence towards the formation of that policy.

The Future Outlook

A few thoughts on what these statistical applications of productivity indices have led me to think lies ahead may not be out of place. Possibilities for the extension of this type of work appear to be extremely great; and the more traditionally intractable the industry, the more stimulating does the prospect become. Models of the problems undertaken by production control are envisaged which would embrace the whole complex manufacturing structure of, say, an integrated steelworks. The type of model discussed here might, through the theory of queues, be extended to deal with the flow problems of hot metal. Steel making itself will perhaps be amenable to inclusion in models as a stochastic process.

More importantly, comprehensive models of the Leontief type might be constructed, and matrices can be visualised incorporating even the huge number of factors involved in a steelworks economy. Thanks to the theories of games and of linear programming, such static models can now be made dynamic. If electronic calculating machines are ultimately to become available at combine and even company level, as seems likely, the routine handling of such matrices will become possible.

Most major works' control problems are problems of balancing factors to maximise profitability. Some form of linear programme using the productivity index as a common measure of effectiveness might provide a means of solving such problems for industries of the most complex type. The possibilities emerge clearly from the Carnegie Institute discussion¹² of their 'nut-mix' problem. The imposing

anthology¹³ of work done in America on activity analysis also provides many pointers to the future.

Advances in the increasingly discussed subject of cybernetics, allied with the complex models mentioned, might result in a fully mechanised form of control based on the technique described here. An electronic machine might be fed with punched-card orders, which would promptly be translated into objective times, weighted by a productivity index, and tabulated as production schedules from a linear programming calculation. If the results of production were returned to the machine, fresh productivities could be assessed, as also the statistical probability of the significance of the latest deviation from any given productivity mean. This probability could then be used automatically to adjust the weighting factor by a feed-back device.

Such conjectures as these seem to be far-fetched at present, although various experiments are actually in hand along these lines. But if, as I believe, the 'technology of control' has been absurdly outstripped by the technology of the controlled situation, they may serve to encourage the development of the relatively elementary sort of control scheme proposed in this paper.

Acknowledgements

I wish to thank Mr David A. Hopkins and my other assistants for their help, and the management of Samuel Fox and Company Limited, Stocksbridge Works, Sheffield, for permission to make public this material.

REFERENCES

- ¹ ANGLO-AMERICAN COUNCIL ON PRODUCTIVITY (1950). 'Assessing progress by indexes', *Productivity Measurement in British Industry*.
- ² FISHER, R. A. and YATES, F. (1953). *Statistical Tables* (4th ed.), Table XII. Oliver and Boyd, London and Edinburgh.
- ³ EISENHART, C. (1947). *Selected Techniques of Statistical Analysis*, Chapter 16. McGraw-Hill Book Co., New York, for Columbia University.
- ⁴ STEVENS, W. L. (1953). 'Tables of the angular transformation', *Biometrika*, 40, 70.
- ⁵ BEER, S. (1953). 'A technique for standardising massed batteries of control charts', *Applied Statistics*, 2, 160.
- ⁶ BAINES, A. H. J. (1942). 'Methods of detecting non-randomness in a given series of observations', Report R/12. Ministry of Supply, Advisory Service on Quality Control, London.
- ⁷ ABRUZZI, A. (1952). *Work Measurement*. p. 43 *et seq.* Columbia University Press New York.
- ⁸ WEILER, H. (1952). 'On the most economical sample size for controlling the mean of a population', *Ann. Math. Statist.*, 23, 247.
- ⁹ COLUMBIA UNIVERSITY STATISTICAL RESEARCH GROUP (1945). 'Differentiation of measured lot from standard (first method)', Report 255 (revised), Section 5.
- ¹⁰ TIPPETT, L. H. C. (1935). 'A snap-reading method of making time studies of machines and operatives in factory surveys', *J. Text. Inst.*, 26, T51.
- ¹¹ BEER, S. (1954). 'Operational research and accounting', *Operat. Res. Quart.*, 5, 1.
- ¹² CHARNES, A., COOPER, W. W., and HENDERSON, A. (1953). *An Introduction to Linear Programming*. John Wiley, New York, and Chapman and Hall, London.
- ¹³ KOOPMANS, T. C. (Editor) (1951). *Activity Analysis of Production and Allocation*. Cowles Commission Monograph No. 13. John Wiley, New York, and Chapman and Hall, London.

A METHOD OF GRAPHING SALES FIGURES FOR EXECUTIVE USE

ALBERT W. SWAN

Courtaulds Ltd.

Mr Swan describes methods that have proved useful in an organisation where there is forward buying and, by way of example, discusses the practical implications of the results of applying the method.

THE method described in the following article was designed to cope with a particular set of problems. They are presented as possibly having a wider application, and in the hope that the basic ideas may be found useful.

In an organisation in which orders vary in size and regularity of arrival it is not easy to present a chart of sales results on a regular, say weekly, basis. If, for instance, most customers place orders at weekly intervals and one big customer places his order once a month, a chart giving the amounts ordered each week would be misleading as showing an apparent slump for three weeks out of four and an apparent boom for the fourth week. Yet executives demand, and reasonably, a chart showing the position each week so that they can plan production accordingly.

A further difficulty in charting sales figures is that of scale. The sales of product A may, on the average, be ten times those of product B. If the results for both are plotted on ordinary graph paper to the same scale, results for product B will crawl along the bottom of the chart and changes will be difficult to detect; while if they are plotted on different scales the visual interpretation may be dangerously wrong.

The first method that comes to mind for use in dealing with results that incorporate two different cycles such as weekly and monthly, is the moving average. In the particular case referred to, a five-week weekly moving average would smooth out the monthly 'hump', at any rate to some extent. The moving average is, however, only an expedient and a somewhat insensitive one, so it became necessary to find a better alternative and this was found in the straightforward cumulative chart. At Courtauld's orders are classified by the month of promised delivery, so that it is possible to construct a cumulative chart on which there are lines for delivery in January, delivery in February, delivery in March, etc., each starting from zero and growing to a cumulative total.

However, one would wish to examine such a chart not only in connection with the ultimate total sales for any given month, but for the general rate of growth at any given moment. At the end of May, for

instance, one would want to know not only the total orders that had been received for delivery in May, but the rate at which orders were coming in for delivery in June, July, etc. 'The forward buying position is a strong indication of customers' confidence or lack of confidence.

It will have been noticed that in the above examination one is mainly interested in rate of growth—in graphical terms, slope. Furthermore, one wants a graph on which a given slope will have the same meaning at whatever level on the graph paper it occurs. The solution then is to use graph paper with a logarithmic vertical scale for the sales results and an arithmetic horizontal scale for time. This graph paper is sensibly known in the USA as ratio paper.

Fig. 1 refers to orders for viscose textile continuous filament yarn, and although the vertical scale has been altered so that no actual Courtauld results are disclosed, it corresponds in general shape to the

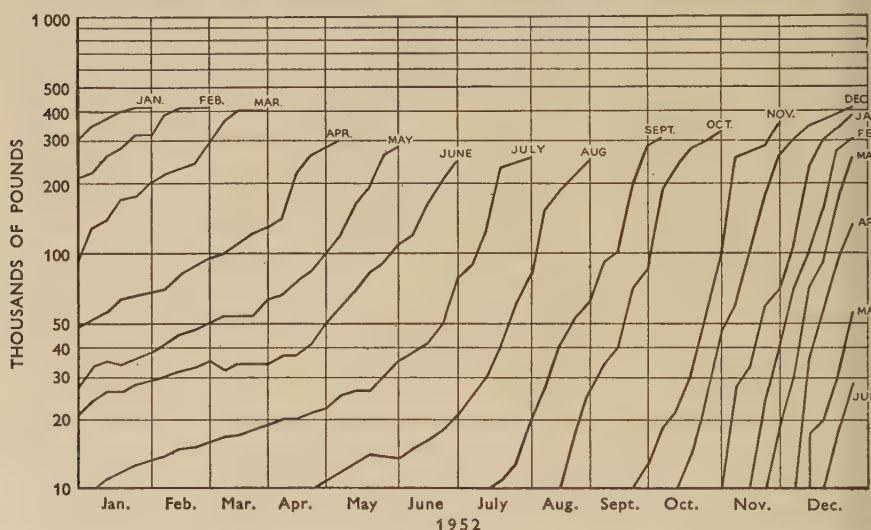


FIG. 1. Cumulative orders for yarn for delivery in current and future months plotted weekly

real chart for 1952. Each line refers to an eventual date of delivery. Thus it will be seen that orders for delivery in July began as far back as January, there being 10 000 pounds on order on 3rd January. These 'July' orders climbed slowly but at a fairly constant rate until about 15th May, by which time there were 25 000 pounds on order. They then took an upward turn, increasing rapidly till the beginning of the month of delivery, July, with 230 000 pounds on order on 10th July, after which date a few orders came in till the end of the month, giving a grand cumulative total of 260 000 pounds.

From the original of Fig. 1 certain main conclusions became evident. The cumulative feature effectively overcomes the difficulty of presenting a weekly statement when orders arrive irregularly, since the cumulative position to date is at once apparent, not only for current but future orders. The use of a logarithmic vertical scale for volume of order

and an arithmetic horizontal scale for time is equally justified, since slope has the same meaning at all levels and visual comparisons of rates of increase are valid. Furthermore, visual comparisons between the progress of orders for the different main types of product, which differ considerably in volume, are also valid.

Let us now look at Fig. 1 as a picture of a year's events. January to July shows an impending and actual slump; the total orders drop from 400 000 to 240 000 pounds per month; the lines climb slowly, showing a slow arrival of orders, and they are far apart, showing timidity in placing future orders. July to December show the astoundingly quick recovery. The total sales per month improved but the most significant point is the completely different typical slope. Every customer was in a hurry to buy; orders poured in, and the lines were almost vertical. By December the volume of orders for future delivery had increased to such an extent that the total orders for three months ahead greatly exceeded 'current' orders.

It has been suggested that the same type of curves would have been shown if there had been a seasonal, summer, decline in orders. This is possibly true, but continuation of the charts since the period illustrated show that the seasonal effect is relatively slight and that the picture is that of a slump and recovery. Similar charts to Fig. 1 were drawn for other types of product. Those for acetate textile yarn were similar in type of curve to those for viscose textile yarn, the typical line being an elongated S. For industrial (tyrecord) and for 'Fibro' cut staple the typical line was found, however, to rise sharply, almost vertically, and flatten out to an almost horizontal line some weeks before the date of promised delivery. These charts thus reflected differing buying habits. The buyers of textile continuous filament yarn bought a long time ahead of delivery, at a steadily increasing rate until just before the promised date of delivery. In the cut staple and tyrecord markets, however, the tendency was more for bulk orders to be placed fairly regularly and at a fairly definite interval before the date of promised delivery.

In the market where the placing of orders for future delivery is characteristic it is important to assess, if possible, the extent of future buying since this fluctuates with the confidence of the buyer in general trade conditions and is, therefore, a valuable trade barometer.

The next step in the study of 'future' buying was to draw Fig. 2 to show at the end of each month the 'current' orders, the orders for '1st month ahead', and for '2nd month ahead'. Fig. 2 was derived from Fig. 1 by joining the required points on Fig. 1. Thus the 'current', '1st month', and '2nd month' lines for the January, February, and March have the values 410, 410, 400; 320, 290, 130; and 200, 96, 66.

Examination of Fig. 2 suggested further steps. The first was the design of a transparent ready reckoner for reading off without calculation (a) the percentage relation of the '1st month' or '2nd month' value to that of the 'current' month (e.g. in January '1st month' and

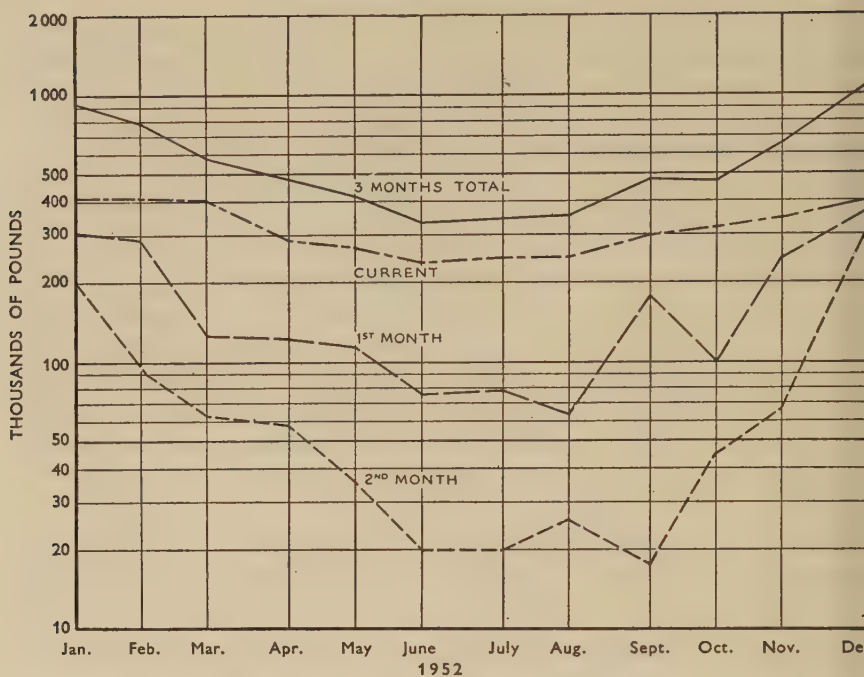


FIG. 2. Orders for current month, one month ahead, two months ahead, and totals for the three months, received up to the end of each month.

1. PERCENTAGES

- Set '100%' on vertical scale over main value.
- Read percentage of second value, etc., on vertical scale.

2. FACTORS

- Set '1' on vertical scale over main value.
- Read 'times multiplied by' for second value, etc., on vertical scale.

3. TRENDS

- Set centre of circle over beginning of line whose rate of increase or decrease per month is required.
- Note where the line cuts the inner circle and project to edge of 'protractor'.
- Read rate of increase or decrease per cent.

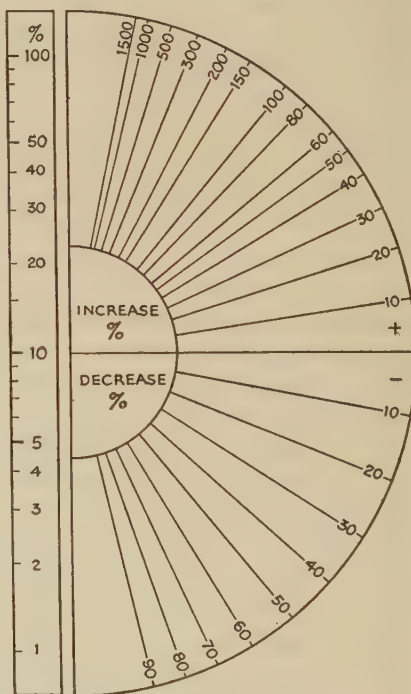


FIG. 3. Ready reckoner for use with charts like that in Fig. 2. Figs. 2 and 3 are not reproduced on the same scale.

'second month' orders are 76 per cent and 49 per cent respectively of 'current' orders), and (b) the rate of increase or decrease of orders per month (e.g. from August to September the '1st month' orders increased at the rate of 150 per cent per month). The ready reckoner is shown in Fig. 3. The vertical scale is a reproduction of two cycles of the vertical logarithmic scale, suitably marked off to read percentages, so that the '100%' mark can be placed on the 'current' line and the appropriate percentage read off for 'first month' and 'second month'. The protractor was laid out to show the percentage increase or decrease per four horizontal time divisions.

So far, so good. The charts fill a need, and are in executive use. It seemed desirable, however, to delve further into the mystery of variations in forward buying by customers. Was it possible to assess in quantitative terms the well-known fact that there is more forward buying in boom and less in slump, and was it possible to link this variability with some factor in the relation of cause and effect? No extravagant claims are made for the results of the investigation now to be described, which was based on the real figures that lie behind Figs. 1 and 2, but they supply food for thought.

If one is examining forward buying on a numerical basis, it is necessary to have some kind of index. A very simple index can be constructed in the following way. Regarding orders for the 'current' month as 100 per cent, find the corresponding percentage values for 'first month' and 'second month' ahead. Add these two to form a 'future' buying index. The reason for taking each relationship, first month to current and second month to current, into account is that the first and second month lines behave independently and their relative position is important. Thus if current sales are 100 000 pounds, first month are 95 000, and second month are 80 000, the inference would be that there is general confidence in future buying. If current and first month continue at about 100 000 and 90 000, but second month gradually drops to 85 000, 80 000, 70 000, etc., the inference would be that buying up to a month ahead is continuing but that buying for longer ahead is declining, and that there is a change in buying habits towards the 'hand to mouth'. This particular change did, in fact, take place at one time subsequent to the period depicted in Figs. 1 and 2 and it was one of considerable importance in the planning of production and control of the level of stock.

To take a particular case from Fig. 2 we find that in January 1952 the actual sales were:

Current	..	..	..	410 000 (= 100 per cent)
Next month	..	..	..	310 000 (= 75.5 per cent)
Two months	..	..	..	200 000 (= 48.8 per cent)

On this basis the Future Index is $75.5 + 48.8 = 124.3$. It will be apparent that the greater the tendency to buy forward, the higher will be the ratio of next month and two months hence sales to those of the

current month. For example, if next month's sales are equal to this month's 100 per cent, the tendency to buy forward is greater than if they were 50 per cent of this month's. Thus the Future Index varies directly as the tendency to place future orders, though the actual relationship with the psychological factors which are behind that tendency are as yet unknown.

The next step was to investigate possible cause and effect relationships for the Future Index. Most of these gave no clear indication, but eventually there was one which, without being stated too dogmatically,

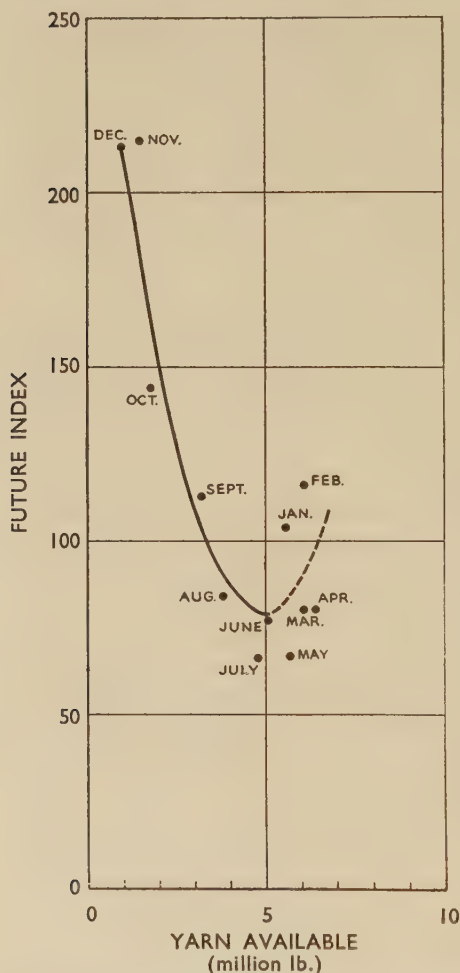


FIG. 4. Relationship for 1952 between the Future Index, which measures the tendency to place orders for future delivery, and the yarn available or the Surplus Stock.

actual Surplus Stocks, and the agreement will be noted. (The figures have been altered from those that actually occurred, but the general effect corresponds to reality.) From the position of the points in relation

was useful. There is a quantity which one can call Surplus Stock, i.e. stock and production not actually booked against orders. For the purpose of this investigation it was defined as actual stock + 3 months' estimated production - 3 months' orders. It was found that when the values of the Future Index were plotted against the first differences of the Surplus Stock, that is the change up or down from one month to another, a regression line was found which was significant within the usual statistical 5 per cent limits. It is interesting but not surprising that it should be the first differences rather than the actual values of Surplus Stock that gave the significant results (the actual values did not). The buyer thinks not so much in terms of an absolute value, but with reference to the change from the last time. It is not so much 'Have our suppliers a good stock available?' as 'Has our suppliers' stock become smaller or larger?'

The results were integrated backwards so to speak, and Fig. 4 shows the theoretical curve setting the Future Index against

to the curve in Fig. 4, it will be seen that those for January to July lead roughly in chronological order down the right-hand part of the parabola to the base. This part, therefore, represents the slump which was severe at first and gradually died away. The points for July onwards to December climb up the left-hand part of the curve and correspond to the period when the slump disappeared and turned into a boom. There may have been an exchange of cause and effect from slump to boom and the right-hand dotted part of the curve may therefore have a different meaning from the left-hand solid part. During a slump the customer does not care how big the supplier's surplus stock may be. As, however, signs of improvement begin to appear, he does begin to wonder about his future supplies at a point corresponding in this case to the bottom of the curve, or in actual date, June and July. As the improvement accelerates and surplus stocks continue to fall the buyer becomes increasingly anxious about his future supplies and increases his forward orders, with a rate which varies inversely as the square of the surplus stock.

It would be dangerous to draw too rigid conclusions from this one example, but it does appear that the control of surplus stock may be more closely linked with buying confidence than has been realised; in the particular case under examination it would seem advisable to hold the level of surplus stock somewhere to the left of the bottom of the curve, high enough to maintain confidence by the buyer and yet low enough to avoid locking up too much capital. This would correspond to a future buying index varying between 80 and 120.

Charts of the type described above are most useful where forward buying is an integral part of the normal commercial practice. Such practice is natural in the textile industry where the actual time of process from raw material to finished goods may extend to several weeks. Experiments are being made in other industries in which the buying patterns differ. It will be interesting to see how they succeed.

AN APPLICATION OF LINEAR PROGRAMMING TO THE SELECTION OF RAW MATERIALS

NOEL WILLIAMS, FIA

Mr Williams illustrates algebraic and arithmetical procedure by working through two simple examples.

IN the ordinary course of business the management is faced with problems of choice. In many cases the problems are fairly simple and can be resolved by common sense and experience, but sometimes they are too complicated for this. Mathematical methods can be used in some of these more complicated problems to work out the best choice to make. Linear Programming is a branch of mathematics which has been designed to deal with problems of choice; it has been developed only since the war and has not, as yet, been used very widely; but interest in the subject is growing.

Linear programming can solve only those types of problem that can be expressed concisely in mathematical notation, where the problem, all facts affecting it, and the solution, can be given numerical values and where the aim is to choose some combination of alternatives out of some greater number of possible alternatives.

A typical problem which can be solved by linear programming is the choice of ingredients to make a product whose quality is expressed as its content of various attributes, where all the possible ingredients contain one or more of these attributes, but where each ingredient has a different analysis. For any set of prices and specified quality of final product it is possible to calculate which ingredients should be purchased, and how much of each, to make the product at the most economical material cost. This problem is dealt with in more detail in the examples below.

A somewhat similar application is the allocation of scarce material between products. The method is likely to prove useful in many problems which involve the selection and mixing of ingredients, and is particularly suitable for the more complicated problems.

Example 1

Suppose that the quality of the product is the percentage content of a certain attribute M, that there are four ingredients from which to choose, and that we wish to obtain a product which contains *at least* 18 per cent of M. The problem is to find out which ingredients to use and how much of each in order to make the product at the minimum cost.

TABLE I
Analysis and cost of ingredients

Ingredient	Percentage Content of M	Cost per Ton, £
A	51	40
B	11	20
C	14	24
D	36	30

In practice an example as simple as this could be worked out without the use of linear programming, but this very simplicity makes it suitable for illustration, as the reader can easily check the solution found below.

If we compare the content of M with the cost per ton we find that, ignoring tonnage, the cost per one per cent of M is as shown in Table II. Ingredient A is thus the cheapest source of M. Ingredient B (at £20 per ton) is the cheapest source of bulk.

TABLE II

Ingredient	Cost per 1% of M, £
A	$40/51 = 0.78$
B	$20/11 = 1.82$
C	$24/14 = 1.71$
D	$30/36 = 0.83$

No one ingredient contains exactly 18 per cent of M, but either A or D could be used on its own. This would be expensive, and the cost could be reduced by mixing B or C with either A or D; considering tonnage and M separately it might be thought that a mixture of A and B to give exactly 18 per cent of M would be our cheapest formula. We can test this hypothesis by working out all possible combinations of ingredients which give a mixture of 18 per cent of M or more and costing them. If we confine ourselves to mixtures giving exactly 18 per cent of M, with only the four ingredients listed above, this would not take long; but in a practical case where we have more ingredients, a more complicated specification of quality, and perhaps limits on the quantities of the various ingredients which could be used, this procedure would not be practicable. We therefore need some systematic procedure which will check our first hypothesis and lead us to the best solution if we fail to obtain it on the first trial.

As a first trial we shall use ingredients A and B. In linear programming we can proceed only by one step at a time, and our first step will be to use the better of these two ingredients, that is A. We compare ingredient A with the product and each ingredient separately by the use of Table III, which shows the original specification and the differences between this specification and one ton of A. As we have equated the tonnage, these differences are confined to the M content, and show the error we make if we try to express the product and each ingredient in terms of A, as we cannot equate both tonnage and M.

TABLE III

Requirement and ingredients

	(i) Original Specification		(ii) In Terms of Ingredient A	
	Tons	M (%)	Tons of A	Surplus of M
	1	18	1	$51 - 18 = 33$
Product	1	18	1	$51 - 18 = 33$
Ingredient A	1	51	1	$51 - 51 = 0$
" B	1	11	1	$51 - 11 = 40$
" C	1	14	1	$51 - 14 = 37$
" D	1	36	1	$51 - 36 = 15$

The next step is to put in ingredient B in such a way that the product and each of the ingredients will be expressed in terms of A and B. In part (ii) of Table III it will be seen that we have a surplus of 40 units of M when B is expressed in terms of A. Call this value of 40 the 'pivot'; divide the column of figures in which the pivot occurs by the value of the pivot, namely 40, and the resulting values, which are shown in column (ii) of Table IV, are the required proportions of B. Subtract these figures from 1 (the first column of figures in part (ii) of Table III) to obtain the required proportions of A.

TABLE IV

Product and ingredients in terms of A and B

	(i) Tons of A		(ii) Tons of B
	0.175	0.825	
Product	0.175	0.825	
Ingredient A	1.000	0.000	
" B	0.000	1.000	
" C	0.075	0.925	
" D	0.625	0.375	

The process of obtaining Table IV can be likened to a mass production procedure for solving a series of simultaneous equations. For example, if x_a and x_b are the quantities of A and B respectively to make one ton of a product containing 18 per cent of M, then

$$x_a + x_b = 1 \text{ (ton)}$$

$$\text{and } 51x_a + 11x_b = 18$$

which on solution gives $x_a = 0.175$ and $x_b = 0.825$. Similarly in expressing ingredient C in terms of A and B we have the equations

$$x_a + x_b = 1 \text{ (ton)}$$

$$\text{and } 51x_a + 11x_b = 14 \text{ (C's content of M)}$$

which gives $x_a = 0.075$ and $x_b = 0.925$.

We now introduce the prices (Table I) to cost the product and to compare the prices of the ingredients with the cost of equivalent mixtures of A and B. These are in Table V.

TABLE V
Comparison of prices with cost of A and B
(based on quantities shown in Table IV)

			Cost of A (@ £40)	Cost of B (@ £20)	Total Cost Z	Price per Ton c	$Z - c$
Product			£ 7.0	£ 16.5	£ 23.5	£ —	£ —
Ingredient A	..		40.0	0.0	40.0	40	0.0
" B	..		0.0	20.0	20.0	20	0.0
" C	..		3.0	18.5	21.5	24	- 2.5
" D	..		25.0	7.5	32.5	30	+ 2.5

If we had been correct in thinking that A and B would give us the cheapest mixture, all the values in the column headed $Z - c$ in Table V would have been zero or negative. The fact that the $Z - c$ for ingredient D is positive shows that we can cheapen our formula by using some of this ingredient. It can be shown that for this problem our cheapest mixture will not contain more than two ingredients. Hence if D is to be used, and as the total quantity of all the ingredients used in the formula amounts to one ton, we shall have to use it in place of A or B or both.

If we imagine our formula to be weighed out in two heaps:

$$\begin{array}{c} \boxed{} \\ \text{A} \end{array} = 0.175 \text{ tons} \qquad \begin{array}{c} \boxed{} \\ \text{B} \end{array} = 0.825 \text{ tons} \dots (H_1)$$

the heap of B can be divided into two smaller heaps in such a way that one of the heaps plus the heap of A has an average M content of 36 per cent (the M content of D). We know from the equation for D that the following have an average M content of 36 per cent:

$$\begin{array}{c} \boxed{} \\ \text{A} \end{array} = 0.625 \text{ tons} \qquad \begin{array}{c} \boxed{} \\ \text{B} \end{array} = 0.375 \text{ tons} \dots (H_2)$$

and if we reduce both in the same proportion, the average will not be affected.

We have to decide how much of D to use, and we are going to use it to replace a mixture of A and B which has an average M content of

36 per cent, thus leaving the M content of the formula unchanged at 18 per cent. Table V shows that D is an economical ingredient, and so we wish to use as much as possible. As we have to take out A and B in the ratio 0.625 : 0.375 we are limited by the amount of A which is already in the formula, hence the maximum amount of D that we can use is $0.175/0.625 = 0.28$. Taking 28 per cent of the heaps shown against H_2 , we get:

$$\begin{array}{l} \boxed{} \\ \text{A} \end{array} = \begin{array}{l} 0.625 \times 0.28 \\ 0.175 \text{ tons} \end{array} \qquad \begin{array}{l} \boxed{} \\ \text{B} \end{array} = \begin{array}{l} 0.375 \times 0.28 \\ 0.105 \text{ tons} \end{array} \quad \dots (H_3)$$

Hence the heaps representing our formula become:

$$\begin{array}{l} \boxed{} \\ \text{A} \end{array} = 0.175 \text{ tons} \quad \begin{array}{l} \boxed{} \\ \text{B} \end{array} = 0.105 \text{ tons} \quad \begin{array}{l} \boxed{} \\ \text{B} \end{array} = \begin{array}{l} 0.825 \\ - 0.105 \\ 0.72 \text{ tons} \end{array} \quad \dots (H_4)$$

We replace the A and the smaller B heap (which together have an average of 36 per cent M) by a single heap of D containing $(0.175 + 0.105) = 0.28$ tons, which also has 36 per cent M content. We have already found that one ton of D is cheaper than 0.625 tons of A plus 0.375 tons of B, the difference in price being $(32.5 - 30) = \pounds 2.5$. As we are substituting 0.28 tons of D for 0.175 tons of A and 0.105 tons of B, the cost of the formula will be reduced by $0.28 \times 2.5 = \pounds 0.7$.

We shall now show how the above method using heaps is handled algebraically, by obtaining this new formula (B = 0.72 tons, D = 0.28 tons) directly from Tables IV and V, and at the same time find out whether it is possible to make any further reduction in the cost. From Table V we see that ingredient D is to be used. In Table IV we use the 'division test' to find out whether it should be substituted for A or for B; the procedure is: divide the figures in the product row by the corresponding figures in the row for ingredient D. This gives us two ratios, $0.175/0.625 = 0.28$ and $0.825/0.375 = 2.2$. We take the *smaller* of these two ratios as the guide and substitute ingredient D for ingredient A. (We take the *smaller* ratio because we cannot remove a larger heap than we have already weighed out. Consider what would happen to A if we used the ratio of the heaps of B against H_1 and H_2 , i.e. $0.825/0.375 = 2.2$. We should have to remove $0.625 \times 2.2 = 1.375$ tons of A from a heap of only 0.175 tons.) The new pivot is 0.625 and the other figure in the row for ingredient D is the 'factor'. Divide all the values in column (i) of Table IV by the pivot; the result is given in column (i) of Table VI. Multiply this new column by the factor (column (ii) of Table VI) and subtract the values from column (ii) of Table IV as shown in column (iii) of Table VI. Column (ii) of Table VI is of n

further use, but we shall require (i) and (iii) for costing the new formula and price comparisons.

TABLE VI
Substitution of D for A

	(i) Tons of D	(ii) (i) $\times$ 0.375	(iii) Tons of B = Table IV col. (ii) - Table VI col. (ii)
Product	0.28	0.105	0.72
Ingredient A	1.60	0.600	- 0.60
" B	0.00	0.000	1.00
" C	0.12	0.045	0.88
" D	1.00	0.375	0.00

TABLE VII
*Comparison of prices with cost of B and D
(based on quantities shown in Table VI)*

	Cost of D (@ £30)	Cost of B (@ £20)	Total Cost $\mathcal{Z}$	Price per Ton c	$\mathcal{Z} - c$
Product ..	£ 8.4	£ 14.4	£ 22.8	£ —	£ —
Ingredient A	48.0	- 12.0	36.0	40	- 4.0
" B	0.0	20.0	20.0	20	0.0
" C	3.6	17.6	21.2	24	- 2.8
" D	30.0	0.0	30.0	30	0.0

The cost of the new formula is £22.8, which is £0.7 less than that of our previous formula. All the values appearing in the $\mathcal{Z} - c$ column are now zero or negative, and we have therefore reached the cheapest formula.

In the above example we have ignored the possibility that a mixture containing more than 18 per cent of M might be cheaper than the result found above. This possibility should, of course, be allowed for, and the method of dealing with it will be shown in the next example.

Example 2

This problem is similar to that in Example 1, except that we have the additional proviso that not more than 20 per cent of the formula is to consist of ingredients B or C or B and C together. The same ingredients are available.

We shall express the requirements and limitations in the form of simultaneous equations. Let x_a , x_b , x_c , and x_d denote the unknown tonnages of ingredients A, B, C, and D to be used. We have the following conditions:

E_1 : The quantity of all the ingredients is to equal 1 ton, therefore $x_a + x_b + x_c + x_d = 1$.

E_2 : The content of M in the product is to be at least 18 per cent therefore $51x_a + 11x_b + 14x_c + 36x_d \geq 18$.

E_3 : As not more than 20 per cent of the formula is to consist of B or C or both, we have $x_b + x_c \leq 0.2$.

E_4 : There is a limitation that no x shall be less than zero, as we cannot have a negative amount of any ingredient. This restriction is not put into the equations but is taken care of by using the smaller ratios brought out by the 'division test' (see page 26).

E_2 and E_3 are inequalities, and we have to change them into equations. We do this by introducing 'difference factors,' denoted by q 's to distinguish them from the ingredients. In E_2 we introduce q_1 , where q_1 is the excess of the M content in the formula over the minimum value of 18. The difference factor q_1 is thus an unknown number ≥ 0 . E_2 becomes $51x_a + 11x_b + 14x_c + 36x_d - q_1 = 18$. In E_3 we introduce q_2 , where q_2 is the difference between the total quantity of ingredients B and C and the upper limit of 0.2. Thus $0 \leq q_2 \leq 0.2$. E_3 becomes $x_b + x_c + q_2 = 0.2$. The q 's are treated for calculation in exactly the same way as the x 's except that they are priced at zero.

Rearranging the terms and writing the equations E_1 , E_2 , and E_3 , we have

$$\begin{aligned} 1 &= x_a + x_b + x_c + x_d \\ 18 &= 51x_a + 11x_b + 14x_c + 36x_d - q_1 \\ 0.2 &= x_b + x_c + q_2 \end{aligned}$$

These equations are written out in matrix form by taking out the x 's and q 's and writing in the coefficients only. For convenience the equations have been written vertically in Table VIII. Before we can

TABLE VIII
Matrix for requirements and ingredients

Requirements ..	(i)	(ii)	(iii)
	1	18	0.2
x_a	1	51	0
x_b	1	11	1
x_c	1	14	1
x_d	1	36	0
q_1	0	- 1	0
q_2	0	0	1

introduce the price we have to obtain a 'feasible' solution which is any solution, regardless of cost, which meets the requirements. A feasible solution to this problem is given in Table X; it will be seen that the costs are those for the ingredients used to meet the requirements, so that until we have met all the requirements with x 's and/or q 's we have no cost with which to value one or more of the columns. Once we have

reached any feasible solution, all later solutions will also comply with the requirements. If we bear in mind the results of Example 1, a feasible solution which consists of one ton of ingredient D appears suitable. Express the requirements and all ingredients in terms of ingredient D (on the lines of Table III); our pivot is thus the 1 in column (i) row x_d of Table VIII and the factors are 36 and 0 in the row x_d . Divide column (i) by the pivot (column (i) Table IX), multiply column (i) of Table IX by the factors, and subtract from columns (ii) and (iii) of Table VIII respectively (see columns (ii) and (iii) of Table IX).

TABLE IX
Requirements and ingredients in terms of D

	(i)	(ii)	(iii)
Ingredients ..	x_d		
Requirements ..	1	— 18	0.2
x_a	1	15	0
x_b	1	— 25	1
x_c	1	— 22	1
x_d	1	0	0
q_1	0	— 1	0
q_2	0	0	1

To complete the feasible solution substitute q_1 in column (ii) and q_2 in column (iii) of Table IX, the pivots being — 1 and 1 respectively and the factors being 0 in both cases. Having obtained a feasible solution (Table X) we cost the formula and all ingredients; x_d is priced at £30 per ton, q_1 and q_2 at 0.

TABLE X
Comparison of prices with cost of D

	(i)	(ii)	(iii)	(iv) Cost of D = Total Cost $\mathcal{Z}$	(v) Price per Ton c	(vi) $\mathcal{Z} - c$
Cost per Ton, £ ..	30	0	0	£	£	£
Ingredients	x_d	q_1	q_2			
Requirements ..	1	18	0.2	30	—	—
x_a	1	— 15	0	30	40	— 10
x_b	1	25	1	30	20	10
x_c	1	22	1	30	24	6
x_d	1	0	0	30	30	0
q_1	0	1	0	0	0	0
q_2	0	0	1	0	0	0

From Table X we see that we have two positive values of $\mathcal{Z} - c$, £10 for B and £6 for C. We take the largest of these two as the guide and

substitute as much of ingredient B into the formula as we can. We apply the division test and get the three ratios $1/1 = 1$, $18/25 = 0.72$, and $0.2/1 = 0.2$. The smallest of these ratios is 0.2, so we substitute x_b for q_2 . Our new pivot is the 1 in column (iii), row x_b of Table X, and the factors are 1 and 25. Divide column (iii) by the pivot and multiply this column by the factors 1 and 25 in turn and subtract from columns (i) and (ii) respectively. The results are given in Table XI.

It should be noted that it is not essential to take the largest positive value of $Z - c$, but it is a good working rule; we should eventually obtain the same final solution if we took some other value of $Z - c$ as our guide; we could, for example, use that value which gives the greatest reduction in the cost of the product, but this rule would not necessarily give us the optimum solution any quicker.

TABLE XI

Requirements and ingredients in terms of B and D and comparison of prices with cost of B and D

	(i)	(ii)	(iii)	(iv) Cost of D	(v) Cost of B	(vi) Total Cost Z	(vii) $Z - c$
Cost per Ton, £	30	0	20	£	£	£	£
Ingredients ..	x_d	q_1	x_b				
Requirements ..	0.8	13	0.2	24	4	28	—
x_a	1	— 15	0	30	0	30	— 10
x_b	0	0	1	0	20	20	0
x_c	0	— 3	1	0	20	20	— 4
x_d	1	0	0	30	0	30	0
q_1	0	1	0	0	0	0	0
q_2	— 1	— 25	1	— 30	20	— 10	— 10

All the values in the $Z - c$ column are now zero or negative, so we have reached the cheapest formula, which is $x_d = 0.8$, $x_b = 0.2$; this formula provides $0.8 \times 36 + 0.2 \times 11 = 31$ per cent of M, which is 13 in excess of the minimum requirements of 18, thus confirming the value of $q_1 = 13$ shown in Table XI. The cost of the formula is £28.

The values in the Z column are useful, as they show the economic cost of the ingredients which we have not used, based on the current prices of the ingredients which we are using, e.g. the price of ingredient A would have to fall by £10 to £30 per ton to make it worth while using it.

Conclusion

The procedure used in Example 2 (known as the Simplex method) can be expanded to larger and more complicated problems, but all the relations must be linear; the price per ton must be independent of the

quantity used. It makes no difference to the final solution what feasible solution is used as a starting point, but it saves time to choose ingredients which are likely to be required in the final solution. It may happen in some cases that zeros will appear in the requirements row, and this may lead to some difficulty. Once a feasible solution has been reached negative values should never appear in the requirements row. Restrictions may take the form of upper or lower or upper and lower limits on the quality requirements and quantities of individual ingredients. These restrictions are expressed as equations on the lines of Example 2.

I have found that problems involving fifteen to twenty ingredients with three to five quality requirements and up to about fifteen quantitative restrictions can be solved in reasonable time on a comptometer or a desk calculating machine and that, with experience, a number of time-saving short-cuts become evident. Once the conditions, requirements, and restrictions have been laid down by the management the answer can be produced as a matter of routine calculation.

BIBLIOGRAPHY

- CHARNES, A., COOPER, W. W., and HENDERSON, A. (1953). *An Introduction to Linear Programming*. Wiley, New York; Chapman and Hall, London.
- LOMAX, K. S. (1952). 'Allocation and programming in modern economics', Manchester Statistical Society.

THE COMPARISON OF MEANS OF INDUSTRIAL TIME SERIES*

GEOFFREY H. JOWETT
University of Sheffield

When a set of observations form a time series it is wrong to use the ordinary random sampling theory for calculating standard errors. Dr Jowett discusses the problem and shows what should be done.

MANY industrial applications of statistics involve the comparison of means calculated not from random samples but from stretches of time series, and require the calculation of the standard errors of such means or their differences. It is a common practice among industrial statisticians to use for this purpose formulae that are really appropriate only to random samples, a practice which can lead to very serious error if carried out uncritically. It is a simple matter to estimate the standard error of the mean of a random sample using no evidence other than that provided by the sample itself; it is much more difficult, and may be impossible, to estimate the standard error of the mean of a stretch of time series without a great deal of supplementary evidence.

My aim in this paper is to clarify the difficulties, to present techniques which may be used in a few simple problems, and to suggest practical applications of these techniques. The computation involved in the techniques is clarified by a worked example, chosen to illustrate as many points as possible.

Time Series and their Variation

Most of the time series arising in statistical investigation in industry are broadly stationary in character in that they oscillate about a fairly stable level with a fairly stable amount of variation. It is with such series that this paper is concerned. There are certain pure 'types' of series, of which series as usually observed may be taken as compounded. The types, illustrated in Fig. 1, are

- (a) the strictly random series;
- (b) the trend-like series with a smooth oscillatory movement;
- (c) series which partake of randomness, but for which the variation within intervals of a given length tends to increase with that length (e.g. series where there are random differences between hours within the day, between days within the week, weeks within the month, and so on); or in other words, successive values tend to be more alike than values far apart.

* This article is based on a paper given at a conference of the Industrial Applications Section of the Royal Statistical Society held at Cranfield in September 1953.

- (d) strictly periodic series (e.g. having a twenty-four hour cycle);
 (e) quasi-periodic series (e.g. business 'cycles') with movements which look like cycles but tend to get out of step as they become more widely separated in time.

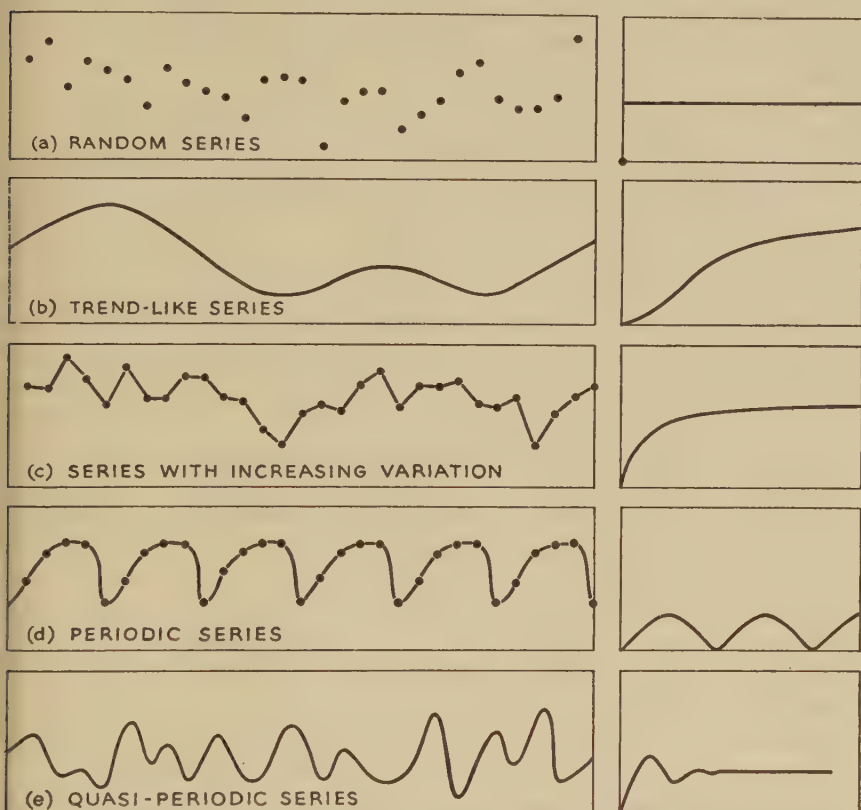


FIG. 1. Types of time series and their serial variation curves. For each right-hand graph, zero for both co-ordinates is at the bottom left corner.

A very important aspect of the variation properties of time series is summarised in the *serial variation function* $\delta(\tau)$ (called the MSSD function in earlier papers^{1,2}). For any given time interval τ , $\delta(\tau)$ is defined as the mean, taken over an infinitely long stretch of the series, or over a population of replications of the series, of half the squared difference between the value of the series at any instant and its value at an instant an interval τ later. When the series is continuous, there is a value of $\delta(\tau)$ for every value of τ —for convenience we define $\delta(-\tau)$ as equal to $\delta(+\tau)$ —and these can be plotted on a graph to give the *serial variation curve*. When the series is discrete, $\delta(\tau)$ exists only when τ is a multiple of the interval w between terms, but in this case it is often convenient to imagine a curve underlying the points plotted on the graph and still speak of the 'serial variation curve' of the series. Serial variation curves for the series from which extracts are shown on the left in Fig. 1 are illustrated (for $\tau > 0$) on the right-hand side. A little

thought is sufficient to make us see roughly why they take the forms they do. (a) In the random series, the separation of two points in time has no effect on the probable magnitude of their difference; the serial variation function is constant (and in fact equal to the variance of the series) except for $\tau = 0$, when it is zero. (b) In the smooth series, the change over a short interval τ is relatively small, and therefore its square is relatively smaller still. Hence the serial variation curve has close contact with the time-axis at $\tau = 0$. (c) In series of this type, the wider the separation of two points in time the greater does the magnitude of the difference between the corresponding values of the series tend to be. Hence as τ increases from zero there is usually a sharp increase (its steepness corresponding to the magnitude of 'local' randomness) which gradually tails off until the curve ultimately flattens out at a value which we may call δ_{∞} . (d) A periodic series has the same value at the same phase of every cycle, and hence the serial variation function is zero for intervals which are multiples of the period of the cycle. The serial variation curve itself oscillates in a periodic manner. (e) For the quasi-periodic series, the serial variation function tends to be periodic for small values of τ , but when τ is large and the cycles get out of step, it settles down to a well-defined value δ_{∞} .

Although fixed periods (associated, for example, with time of day, day of the week, and so on) are common in industrial time series, the strictly periodic component is easy to remove and analyse separately, and the quasi-periodic series is not common in occurrence. Hence in this paper for purposes of exposition it will be assumed that these components are absent from the series under investigation, though the methods are valid when they can be applied to such series. I have examined about thirty time series from various industrial sources, and almost all behave as if made up of elements of types (b) and (c).

In practice, infinitely long series do not exist, so that the values of the serial variation function are hypothetical, like the mean and standard deviation of a hypothetical frequency distribution. The value δ_{∞} , if required, has to be guessed from the direction in which the serial variation curve is going (which is a perfectly reasonable thing to do if the curve has flattened out within the range of experience); it is a hypothetical value, corresponding to the hypothetical maintenance of the *status quo*; it is actually the theoretical variance of an infinite length of series, and though the observed variance of the series is a help in locating it, the latter is apt to be very biased on the low side if the flattening out takes a time which is not a small fraction of the length of the series. If, using the more conventional types of analysis, we assume that a time series is random, we are really carrying out a procedure which is tantamount to guessing δ_{∞} .

The Fitting of Serial Variation Curves

A discussion of the problem of fitting serial variation curves arithmetically lies beyond the scope of this paper. The problem has been

only partly solved, and at the present stage I maintain that a graphical procedure is most to be recommended. The raw material, computed from an observed series, takes the form of serial variation statistics,

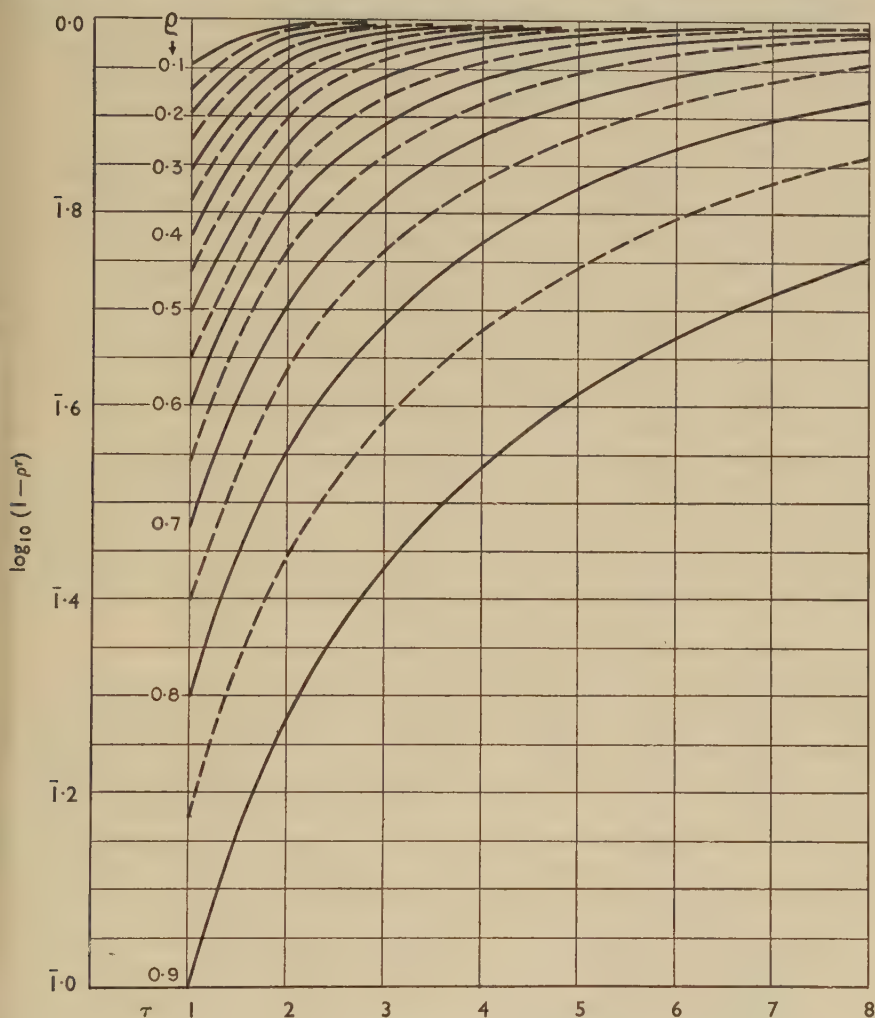


FIG. 2. Chart of $\log_{10} (1 - \rho^\tau)$.

$d(\tau)$. These are worked out by the method which is illustrated below in an example much shorter than would normally be used:

Series (half-hour intervals): 1 3 4 4 5 3 0 2

$$d(0) = 0$$

$$d\left(\frac{1}{2} \text{ hour}\right) = \frac{1}{2}[(1-3)^2 + (3-4)^2 + (4-4)^2 + \dots + (0-2)^2]/7 = 3.3$$

$$d(1 \text{ hour}) = \frac{1}{2}[(1-4)^2 + (3-4)^2 + (4-5)^2 + \dots + (3-2)^2]/6 = 6.3$$

If the series is continuous, the calculation may be carried out on a set

of discrete values taken from it. Missing values merely cause the loss of the terms which involve them. A quick graphical method of working out these statistics has been described in a paper by Mr J. F. Scott and myself²; so far we have usually found it expedient to calculate $d(w)$, $d(2w)$, $d(3w)$, $d(4w)$, $d(5w)$, $d(10w)$, $d(15w)$, $d(25w)$ and possibly $d(50w)$, which give a good indication of the form of the underlying curve.

These values may be plotted on a graph called a *variogram*, and the curve sketched in by eye, with due regard to the following considerations: (i) the sampling errors of the statistics tend to increase rather rapidly with the interval of separation; those corresponding to a long interval are apt to be rather erratic; (ii) the statistics do not vary independently, but tend to cling to their neighbours.

In Fig. 2 are shown curves given by the function $\log_{10}(1 - \rho^r)$, which have been found computationally useful in describing the form of the curve near $\tau = 0$; the logarithms of the first few serial variation statistics are plotted on tracing paper (taking a unit of 0.1 in the logarithm = $\frac{1}{2}$ in.) and superimposed on these curves in turn until a visually satisfactory fit is obtained; the chosen curve is then traced and the fitted serial variation curve obtained by looking up antilogarithms of its ordinates or from the formula itself, and smoothly modifying its shape to fit the later serial variation statistics. These mathematical curves have no magical prognostic value for estimating δ_∞ .

Statistical Analysis with the Aid of the Serial Variation Curve

Methods of analysis involving the use of the serial variation curves fall mainly into two groups. For the first group, knowledge of the behaviour of the series in the large, and therefore an estimate of the value of δ_∞ , is required; for the second group it is only the nature of the short-term or *local* variation which matters, and the value of δ_∞ is not required.

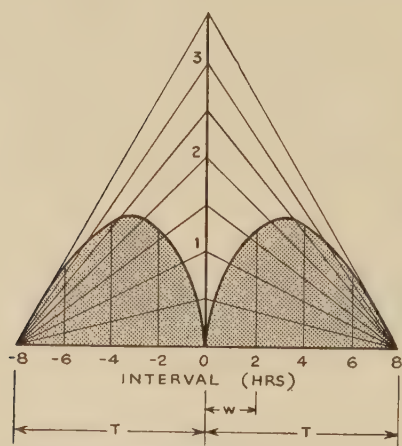


FIG. 3. Serial variation function $\delta(\tau)$ plotted on 'triangular' graph paper.

On this the vertical grid lines are equally spaced as usual, but instead of horizontal grid lines there are lines which slope downwards from

Although in practice it is usually most convenient to use arithmetical methods of computation (this is in fact done in Appendix I), the formulae required are most easily comprehended and remembered when interpreted graphically, and this will be done here. For this purpose it is convenient to use the idea of a special kind of graph paper designed as shown in Fig. 3.

a uniform scale along the central vertical grid line to a point of convergence on either side. I shall refer to this paper as 'triangular graph paper', though the reader must bear in mind that it is merely a useful expository device, and quite different from the standard triangular graph paper that one can buy from stationers. Plotting a curve on this paper has the effect of reducing its ordinates in proportion to their distance from the centre.

The Standard Error of a Mean

If the series is continuous, $\text{Var}(\bar{x}_T)$, the square of the standard error of the mean $\bar{x}_T$ of a stretch of continuous series of length T , may be determined as follows. The serial variation curve is plotted on triangular graph paper as shown in Fig. 3, taking the semi-base as T . Next the area under the curve is determined. Finally the sampling variance of $\bar{x}_T$ is calculated from the formula

$$\text{Var}(\bar{x}_T) = \delta_\infty - (\text{area under curve}/T) \quad \dots(1)$$

The error variance $\text{Var}(\bar{x}_n)$ of the mean $\bar{x}_n$ of n terms of the series separated by a fixed interval w , where $nw = T$, is determined by measuring and adding the $2n - 1$ ordinates of the curve, plotted on the triangular paper as before, at the positions $0, \pm w, \pm 2w, \dots, \pm (n - 1)w$, and finally substituting the total in the formula

$$\text{Var}(\bar{x}_n) = \delta_\infty - (\text{total of ordinates}/n) \quad \dots(2)$$

The standard errors of these means are predictions of the magnitude of their variation about the supposedly stable mean of the series as a whole, and it is impossible to dispense with the need for an estimate of δ_∞ in calculating them. Inability to estimate δ_∞ corresponds to inability to predict the future behaviour of the series, and automatically implies inability to predict how much the mean of a sample stretch may be expected to vary from stretch to stretch. If, with the aid of serial variation statistics computed from an observed stretch of series, we manage to extrapolate to δ_∞ on the reasonable grounds that the serial variation curve appears to settle down to some fixed level and there is no obvious reason why it should increase beyond that level, we ought properly to attach to our standard error estimates the qualification 'if things go on as they appear to be doing at present'. There is no way of avoiding this difficulty; any technique which appears to do so is sure to involve some hidden assumption which is equivalent to a guess at δ_∞ . Sometimes, but only under favourable conditions, it is possible to use arithmetical methods of estimating δ_∞ and to assign standard errors to the figures so obtained. A discussion of this problem, however, lies beyond the scope of this paper.

For some purposes, such as systematic sampling, the standard error of $\bar{x}_n$ about $\bar{x}_T$ may be required. This may be obtained from the formula

$$\text{Var}(\bar{x}_n - \bar{x}_T) = \text{area under curve}/T - \text{total of ordinates}/n \quad \dots(3)$$

Here it is assumed that the first observation is located at random in the first n th part of the interval T .

From formula (3) δ_∞ is absent, and the technique of estimating $\text{Var}(\bar{x}_n - \bar{x}_T)$ therefore falls into the second of the groups mentioned above. The situation is even better than it appears at first sight, since except in an interval just round $\tau = 0$ the area under the curve is approximately proportional to the sum of ordinates, and hence the form of the curve, so long as it is smooth enough for this to happen, is unimportant. Hence this standard error may be estimated from a knowledge of local properties alone.

Example. In Fig. 4 is shown part of a continuous record of the variation of the percentage of nitrogen in the gas from a blast furnace. Assuming that this is stable, we fit a serial variation curve as shown and obtain an estimate of 5.28 for δ_∞ . It is this curve which has been plotted on triangular graph paper in Fig. 3, taking $T = 8$ hours (i.e. one shift). The area under the curve is 14.51 units. Hence, using (1) we find that the standard error of the mean of a continuous eight-hour record is $\sqrt{(5.28 - 14.51/8)}$, or 1.86.

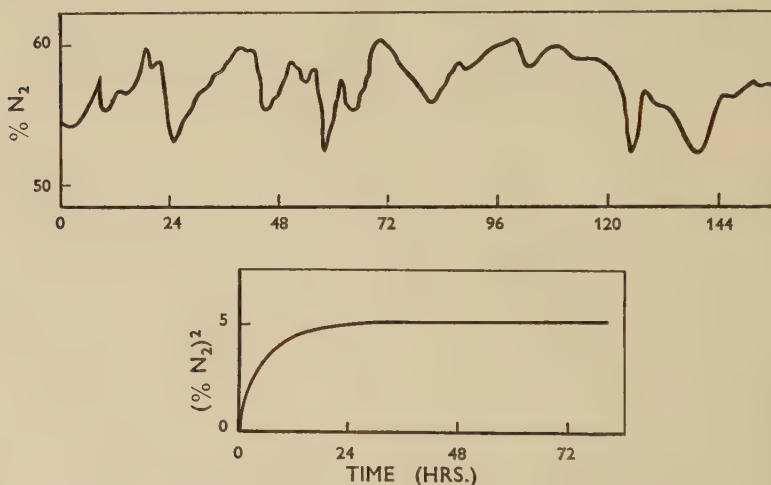


FIG. 4. Graph of percentage nitrogen in furnace gas and its serial variation curve.

If we take four single observations at two-hourly intervals throughout the shift, we have $n = 4$, $w = 2$. The sum of ordinates at $0, \pm 2, \pm 4, \pm 6$ is then 6.76, and from (2) the standard error of the mean of four two-hourly readings proves to be $\sqrt{(5.28 - 6.76/4)}$, or 1.89.

The standard error of the mean of four two-hourly readings about the mean of a continuous record (with the first reading located at random in the first two hours) is calculated, using (3), as $\sqrt{(14.51/8 - 6.76/4)}$, or 0.35.

Applications. I have already discussed applications of formula (3).¹ Formulae (1) and (2) might find application in the solution of problems such as the following.

(i) Fixing limits on a control chart for means of small samples taken systematically over an interval. In these cases the value of δ_∞ used would be based on experience of the variation which could be attained in the absence of assignable (and potentially removable) causes of variation.

(ii) Assessing the standard error of the mean of a long stretch of data from a stable process where δ_∞ appears to be effectively attained by the time τ has reached a size which is only a small fraction of the total length of the observed stretch; it is thus capable of being estimated from the series itself. This standard error might, for example, be used in comparing the mean percentage of nitrogen in furnace gas for furnaces operating under different conditions.

The Error Variance of a Difference between Means

The sampling variance of the difference between means of observations from two independent time series is the sum of the sampling variances of the means themselves, just as for random samples. This is not true for means of observations from the same series.

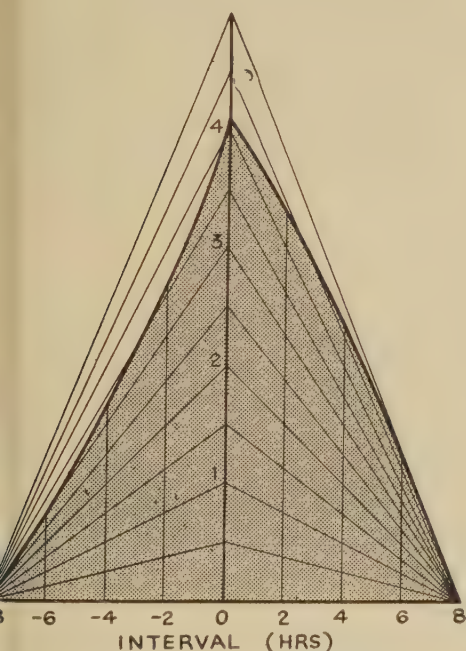


FIG. 5. Displaced serial variation function $\delta(\tau + 11)$ plotted on 'triangular' graph paper.

Suppose two sets of n observations from the same time series are denoted by $x_0, x_1, \dots, x_n, x'_0, x'_1, \dots, x'_n$ with respective means $\bar{x}_n, \bar{x}'_n$; furthermore, suppose that they are each evenly spaced at intervals w , and that x_0, x'_0 are separated by an interval $k w$, where k need not be an integer.

The serial variation curve is plotted on the triangular graph paper as before, taking the semi-base as $n w$, with its centre at the centre of the base of the triangle. The sum of ordinates is calculated as before. The curve is then displaced through a distance $k w$ to the left and plotted again, as in Fig. 5, to give a new sum of ordinates at

$\pm w, \pm 2w, \dots, \pm n w$. The standard error of the difference between the means is then obtained from the formula

$$\text{S.E.}(\bar{x}_n - \bar{x}'_n) = \frac{2}{n} \left[\begin{array}{l} \text{sum of ordinates of displaced curve} \\ \text{minus sum of ordinates of central curve} \end{array} \right] \dots (4)$$

It will be observed that it is necessary to estimate $\delta(\tau)$ as far, at most,

as $\tau = (n + k)w$. The method may be adapted in the obvious way to cover the case of means of sections of continuous record, using areas instead of ordinate sums.

If k is small compared with n (e.g. $k < 1$) so that the two groups of observations overlap (as in the case of interpenetrating samples), it is more convenient to use a different though arithmetically equivalent method.

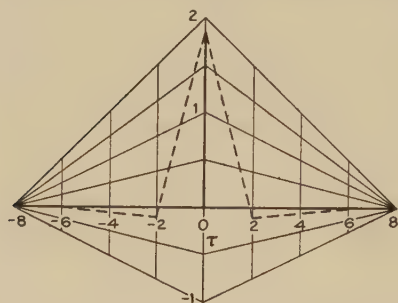


FIG. 6. Values of $\Delta_1''\delta(\tau)$ plotted on 'triangular' graph paper.

values in the neighbourhood of $s = 0$. All the necessary information is often provided by $\delta(\tau)$ for quite small values of τ . The values of the second difference for $\tau = 0, \pm w, \pm 2w, \dots, \pm nw$ are then plotted on the triangular graph paper and the resulting ordinates measured and added, with due regard to sign. The standard error of $\bar{x}_n - \bar{x}_n'$ is then obtained from the formula

$$\text{Var}(\bar{x}_n - \bar{x}_n') = \text{total of ordinates}/n \quad \dots (5)$$

Example. The standard error of the difference between the mean of four two-hourly observations of the percentage of nitrogen in furnace gas and the mean of four two-hourly observations taken eleven hours later (i.e. with the first readings in the two groups separated by eleven hours) is obtained as described above by using formula (4) and taking $w = 2$ hours and $kw = 11$ hours. It is this example which is illustrated in Fig. 5. The sum of ordinates of the central curve (shown in Fig. 4) is 6.76 as before, and that of the ordinates of the displaced curve is 16.27; hence the standard error of the difference between means is $\sqrt{2}\sqrt{(16.27/4 - 6.76/4)}$, or 2.18.

The standard error of the difference between the means of two continuous eight-hour periods separated by a three-hour interval is obtained from the corresponding areas, which are respectively 14.51 and 32.46; its value is $\sqrt{2}\sqrt{(32.46/8 - 14.51/8)}$, or 2.12.

If $\bar{x}_4$ is the mean of four observations at two-hourly intervals and $\bar{x}$ the mean of an interpenetrating sample of four observations so situated that $kw = 1$, we make use of the following values:

$$\begin{array}{cccc} \tau \text{ (hours):} & 0 & \pm 2 & \pm 4 & \pm 6 \\ \Delta_1''\delta(\tau) \equiv \delta(\tau - 1) - 2\delta(\tau) + \delta(\tau + 1): & 1.80 & -0.13 & -0.09 & -0.06 \end{array}$$

The figures in the last line are plotted as in Fig. 6, and the heights of the plotted points above the base line added, with due regard to sign, to give a total 1.49. The standard error of $\bar{x}_4 - \bar{x}_4'$ is then $\sqrt{(1.49/4)}$, or 0.61.

Applications. (i) In the systematic sampling of wagons of coal, an occasional interpenetrating sample is taken as a check on the stability of the variation; if this changes, the accuracy of the systematic sample ceases to be as predicted, and it may become necessary to revise the sampling procedure if the accuracy required is to be maintained. Valid control limits for the permissible difference between sample and interpenetrating sample means may be calculated using a standard error figure derived as above.

(ii) If n is so large that a suitable and sufficiently reliable estimate of $\delta_{(n+k)w}$ may be obtained from within-sample comparisons, the standard error calculated as described above may be used to test the significance of apparent differences in level between two sections of a time series, e.g. before and after some change in practice.

(iii) A systematic experimental design in which treatments occur in the order $T_1 T_2 T_3 T_1 T_2 T_3 T_1 T_2 T_3 \dots$ (e.g. alternations of shift) furnishes three mutually interpenetrating sets of observations. The standard error of the difference between two treatment means should therefore be estimated by the method described above. In the estimation of $\delta(\tau)$ it is necessary to remove the treatment effect by subtracting the appropriate treatment mean from each observation before the serial variation statistics are calculated.

Comparison with Random Sample Methods

The methods of determining standard errors given above are correct for stationary time series; those which involve only local properties are also correct for series which are not strictly stationary, provided the local variational properties are stable enough to give $\delta(\tau)$ some meaning for the *required* values of τ . They are not, however, particularly easy to apply because of the need to estimate the serial variation curve. Hence it is instructive to see what kind of error would be committed by analysing the series as if they were random, using the ordinary methods.

If the series of the example were assumed totally random, its variance being 5.28, so that the sample of four two-hourly readings were regarded as a random sample of four, the standard error of its mean would be estimated as $\sqrt{(5.28/4)}$ or 1.15. This is only about three-fifths of the correct value, and the discrepancy increases with sample size; for eight one-hourly readings, the corresponding fraction would be less than one-half.

If the series of the example were assumed random within shifts, with a within-shift variance which may be shown to be 1.81, the standard error of the mean of four two-hourly readings about the mean of the

corresponding eight hours continuous record would be estimated as $\sqrt{(1.81/4)}$ or 0.67; this is about twice the correct value.

In general, analysis using random assumptions tends to yield serious underestimates of error when applied to problems involving the behaviour of series in the large; for problems involving local properties, the direct opposite is the case. Accordingly it seems desirable that investigators concerned with industrial time series should bear the methods of this paper in mind, even if they ultimately resolve to use the classical random-sample techniques in some form or other.

Acknowledgements are due to Miss W. M. Gibson who was responsible for many of the calculations. The data used in the example were drawn from Special Report No. 18 of the Iron and Steel Institute.

APPENDIX I: NUMERICAL ANALYSIS OF THE EXAMPLES

Seventy readings at intervals of one-tenth of a day were taken from the continuous record. These are given in Table I. Serial variation

TABLE I

1-14	15-28	Day No. 29-42	43-56	57-70
55.1	56.9	59.2	59.2	54.4
53.1	57.5	60.0	57.9	52.9
53.8	59.2	59.2	59.0	51.8
57.7	58.9	58.5	59.5	53.5
54.5	57.0	57.5	59.2	56.0
56.4	54.7	56.1	58.3	56.1
55.3	56.8	54.8	58.3	56.3
59.7	58.8	56.9	58.5	57.0
57.7	56.9	58.5	58.2	56.1
57.5	58.1	57.6	56.4	56.5
53.8	51.5	58.5	51.8	58.6
52.6	56.2	59.2	56.1	59.2
55.0	56.8	59.4	55.2	58.5
56.1	54.6	60.0	55.0	57.7

statistics were computed for various intervals as shown in Table II, and those for lags 1-5 were roughly graduated by the function $4.5(1 - 0.8^\tau)$ where τ was measured in hours. The fitted function, which gives an

TABLE II

Interval days	τ hours	Serial Variation Statistic	$4.5(1 - 0.8^\tau)$
0.1	2.4	2.04	1.97
0.2	4.8	2.95	3.08
0.3	7.2	3.69	3.71
0.4	9.6	4.09	4.07
0.5	12	4.10	4.27
1.0	24	5.07	
1.5	36	5.59	
2.5	60	4.11	

unusually good fit, was used to give values of $\delta(\tau)$ up to 10 hours; beyond this it flattened out too rapidly and was modified to give a reasonable fit to the remaining points by freehand sketching, giving the remaining values (up to $\tau = 19$ hours) required in analysis in Table III. There is, of course, no virtue in this function apart from computational convenience.

The variogram suggested a serial variation curve flattening out within an interval of two days or so, and the variance, which was 5.28, seemed a reasonable limit to use as δ_{∞} . This is not always the case.

TABLE III

Interval hours	$\delta(\tau)$	$(1 - \tau /8)\delta(\tau)$	$\delta(\tau + 11)$	$(1 - \tau /8)\delta(\tau + 11)$	$\Delta_1''\delta(\tau)$	$(1 - \tau /8)\Delta_1''\delta(\tau)$
8	3.75	0	2.21	0		
7	3.56	0.45	2.66	0.33		
6	3.30	0.83	3.02	0.76	- 0.02	- 0.005
5	3.02	1.13	3.30	1.24		
4	2.66	1.33	3.56	1.78	- 0.09	- 0.05
3	2.21	1.38	3.75	2.34		
2	1.62	1.22	3.90	2.93	- 0.13	- 0.10
1	0.90	0.79	4.02	3.52		
0	0.00	0.00	4.13	4.13	+ 1.80	+ 1.80
1	0.90	0.79	4.24	3.71		
2	1.62	1.22	4.33	3.24	- 0.13	- 0.10
3	2.21	1.38	4.43	2.77		
4	2.66	1.33	4.50	2.25	- 0.09	- 0.05
5	3.02	1.13	4.59	1.72		
6	3.30	0.83	4.66	1.17	- 0.02	- 0.005
7	3.56	0.45	4.73	0.59		
8	3.75	0	4.80	0		

From the fitted values the following results were calculated:

Area under curve $(1 - |\tau|/8)\delta(\tau) = 14.51$

Sum of ordinates of $(1 - |\tau|/8)\delta(\tau)$ at 0, ± 2 , ± 4 , $\pm 6 = 6.76$

Sum of ordinates of $(1 - |\tau|/8)\delta(\tau)$ at 0, ± 1 , ± 2 , . . . , $\pm 7 = 14.26$

Area under curve $(1 - |\tau|/8)\delta(\tau + 11) = 32.46$

Sum of ordinates of $(1 - |\tau|/8)\delta(\tau + 11)$ at 0, ± 2 , ± 4 , $\pm 6 = 16.27$

Sum of ordinates of $(1 - |\tau|/8)\Delta_1''\delta(\tau)$ at 0, ± 2 , ± 4 , $\pm 6 = 1.49$

The area under the curves was calculated by dividing the range into four panels and applying to each the integration formula (applicable to functions with continuous derivatives and negligible sixth differences):

$$\frac{90}{c} \int_a^{a+c} f(x) dx = 7f(a) + 32f(a + \frac{1}{4}c) + 12f(a + \frac{1}{2}c) + 32f(a + \frac{3}{4}c) + 7f(a + c)$$

The final computations for the standard errors are in Table IV.

TABLE IV

	Sampling Variance	Standard Error
(i) Mean of shift's continuous record	$5.28 - 14.51/8 = 3.47$	1.86
(ii) Mean of four two-hourly readings	$5.28 - 6.76/4 = 3.59$	1.89
(iii) Mean of four two-hourly readings about mean of shift, with first reading located at random within the first two hours	$14.51/8 - 6.76/4 = 0.12$	0.35
(iv) Mean of eight one-hourly readings	$5.28 - 14.26/8 = 3.50$	1.87
(v) Difference between means of two eight hour continuous records with starting points eleven hours apart	$2[32.46/8 - 14.51/8] = 4.50$	2.12
(vi) Difference between means of two groups of four two-hourly readings, with the first readings separated by eleven hours	$2[16.27/4 - 6.76/4] = 4.76$	2.18
(vii) Difference between means of two interpenetrating samples of two-hourly readings with phase difference of one hour	$1.49/4 = 0.37$	0.61

APPENDIX II: DERIVATION OF FORMULAE

The formulae (1)–(5) are all special cases of one general formula which is not difficult to derive. Let $x(a)$, $x(a + w)$, ..., $x(a + (n - 1)w)$ and $x(a')$, $x(a' + w')$, ..., $x(a' + (n' - 1)w')$ be two sets of observations from a time series and let their means be $\bar{x}$, $\bar{x}'$ respectively. Then

$$\text{Var}(\bar{x} - \bar{x}') = E(\bar{x} - \bar{x}')^2$$

$$\begin{aligned}
 &= E \left[\frac{1}{nn'} \sum_{s=0}^{n-1} \sum_{s'=0}^{n'-1} \{x(a + sw) - x(a' + s'w')\} \right]^2 \\
 &= \frac{1}{(nn')^2} E \left[\sum_{s=0}^{n-1} \sum_{s'=0}^{n'-1} \sum_{t=0}^{n-1} \sum_{t'=0}^{n'-1} \{x(a + sw)x(a + tw) \right. \\
 &\quad \left. + x(a' + s'w')x(a' + t'w') \right. \\
 &\quad \left. - x(a + sw)x(a' + t'w') \right. \\
 &\quad \left. - x(a' + s'w')x(a + tw)\} \right] \dots (6)
 \end{aligned}$$

$$\begin{aligned}
 \text{Now } x(a + sw)x(a + tw) &= \frac{1}{2}[\{x(a + sw)\}^2 + \{x(a + tw)\}^2] \\
 &\quad - \frac{1}{2}[x(a + sw) - x(a + tw)]^2
 \end{aligned}$$

Using this and three similar identities in (6), we have

$$\begin{aligned}\text{Var}(\bar{x} - \bar{x}') &= \frac{1}{(nn')^2} E \left[\sum_{s, s', t, t'} \left[-\frac{1}{2} \{x(a + sw) - x(a + tw)\}^2 \right. \right. \\ &\quad - \frac{1}{2} \{x(a' + s'w') - x(a' + t'w')\}^2 \\ &\quad + \frac{1}{2} \{x(a + sw) - x(a' + t'w')\}^2 \\ &\quad \left. \left. + \frac{1}{2} \{x(a' + s'w') - x(a + tw)\}^2 \right] \right] \\ &= \frac{1}{(nn')^2} \sum_{s, s', t, t'} \{ \delta(sw - t'w') + \delta(s'w' - tw) \\ &\quad - \delta((s - t)w) - \delta((s' - t')w') \}\end{aligned}$$

Collecting together like terms, we may reduce this to the simpler form:

$$\begin{aligned}\text{Var}(\bar{x} - \bar{x}') &= \frac{2}{nn'} \sum_{s=0}^{n-1} \sum_{s'=0}^{n'-1} \delta(sw - s'w') - \frac{1}{n} \sum_{s=-(n-1)}^{+(n-1)} \left(1 - \frac{|s|}{n}\right) \delta(sw) \\ &\quad - \frac{1}{n'} \sum_{s=-(n'-1)}^{+(n'-1)} \left(1 - \frac{|s|}{n'}\right) \delta(s'w') \quad \dots (7)\end{aligned}$$

This general formula may be applied to any series for which the trial variation parameters involved exist. It is thus not always necessary for the series to be strictly stationary; a kind of *local* stationariness may be all that is required. The theoretical implications of this will be discussed in a later paper.

If in (7) we put $w' = w$, $n' = n$ and $a' = a + kw$, the first and third summations become identical and the formula reduces to (4), or with some rearrangement to (5), which is generalised to the continuous case by letting w tend to zero and nw tend to T .

If the series is stationary with mean μ , $\bar{x}' \rightarrow \mu$ as $n' \rightarrow \infty$ and

$$\begin{aligned}\frac{1}{n} \sum_{s=-(n'-1)}^{+(n'-1)} \left(1 - \frac{|s|}{n}\right) \delta(s'w') &\rightarrow \delta(\infty) \\ \frac{2}{nn'} \sum_{s=0}^{n-1} \sum_{s'=0}^{n'-1} \delta(sw - s'w') &\rightarrow \delta(\infty)\end{aligned}$$

respective of the value of w' . Also $\text{Var}(\bar{x} - \bar{x}') \rightarrow \text{Var}(\bar{x})$. In this case (7) reduces to (2). Formula (1) is a limiting case, as before.

A formula equivalent to (2) has been given in terms of autocorrelations by Bayley and Hammersley,⁴ who discussed its application to

autoregressive series. It is my belief, however, that the use of serial variation coefficients rather than autocorrelations makes application simpler, besides extending the validity of the formulae.

REFERENCES

- ¹ HEBDEN, J. and JOWETT, G. H. (1952). 'The accuracy of sampling coal', *Applied Statistics*, **1**, 179.
- ² JOWETT, G. H. and SCOTT, J. F. (1953). 'Simple graphical techniques for calculating serial and spatial correlations, and mean semi-squared difference', *J. R. Statist. Soc.*, **B**, **15**, 81.
- ³ JOWETT, G. H. (1952). 'The accuracy of systematic sampling from conveyor belts', *Applied Statistics*, **1**, 50.
- ⁴ BAYLEY, G. V. and HAMMERSLEY, J. M. (1946). 'The "effective" number of independent observations in an autocorrelated time series', *J. R. Statist. Soc. Suppl.*, **8**, 184.

SOME MAKESHIFT METHODS OF ANALYSIS APPLIED TO COMPLEX EXPERIMENTAL RESULTS*

G. P. DAVIES and G. W. SEARS
The Shell Petroleum Company Limited

After an experiment on the effect of different lubricating oils on the fuel consumption of motor buses, Mr Davies and Mr Sears were confronted with results that did not conform to any tidy text-book pattern. This article describes how orthodox methods of analysis were adapted and used to lead to useful conclusions.

Introduction

The work described in this paper relates to a small part of a general investigation aimed at evaluating the improvement in fuel consumption that could be expected from running vehicles on lubricating oils of lower viscosity than has hitherto been general practice. The experimental design was one of the simplest and merely involved the recording of monthly fuel consumption data on balanced numbers of vehicles of two engine makes, operating in a factorial arrangement on lubricating oils of standard and low viscosity. On the face of it the analysis of the experimental data should have been perfectly straightforward and well within the grasp of anyone with a fair understanding of one of the several practical manuals, now available to the experimentalist, that explain the mechanics of the analysis of variance without giving any detailed explanation of the principles involved or of the conditions necessary for the validity of the estimates and tests of significance. In fact very few of these conditions are satisfied by the data obtained and this paper describes the methods employed to overcome the difficulties arising from this fact. The authors make no claim for the exact validity of these methods, but are satisfied that, in the absence of any known exact methods, they form a reasonable working basis when coupled with an appropriate degree of caution in the interpretation of their results.

The Experimental Data

Twelve double-decker buses were available for this work, all operating a regular transport service under essentially the same service conditions in respect of loading, type of driving, etc. Eight of these vehicles were of one engine make and four of another and it was originally planned to put half of each make on the low viscosity oil leaving the

* This article is based on a paper given at a conference of the Industrial Applications Section of the Royal Statistical Society held at Cranfield in September 1953.

other half on the standard grade. Unfortunately it was not possible to bring all the vehicles on to test at the same time and, in addition to this, three of the vehicles had to be taken out of service for maintenance during the period of the tests. The available data are given in full in Table I and it will be seen that apart from the general unbalance due to the fact that certain of the vehicles came on to test at a later date than others, there are four missing values due to those vehicles which had to be taken out of service for maintenance.

TABLE I

Fuel oil consumption rates, gallons per 1000 miles

Lubricating Oil	Low Viscosity						Ordinary Grade					
Engine Make	A				B		A				B	
Vehicle No.	1	2	3	4	5	6	7	8	9	10	11	12
<i>Summer months</i>												
May		114	111				113		114		112	
June		108	109		107		117	115	124		111	
July		110	112		106	105	116	118	117		112	
August		110	110		116	98	117	111	114		110	
September ..		—	114		111	102	123	117	120		115	
October ..	118	110	115		102	—	126	119	127		114	
<i>Winter months</i>												
November ..	113	112	112	118	115	—	125	121	123	121	112	
December ..	113	117	114	115	112	104	128	125	131	122	116	117
January ..	116	109	114	114	113	104	123	123	126	118	115	119
February ..	108	108	109	110	102	101	123	119	128	120	110	114
March ..	115	116	112	110	113	102	119	120	127	119	116	108
April ..	110	112	112	112	111	110	—	121	119	122	112	116

There was some reason for believing that the saving in fuel consumption might depend both on passenger loading and on ambient temperatures. No details of these two factors were available, but since they are both to a certain extent subject to seasonal variations, the year was divided into two parts, May to October being designated Summer and November to April Winter.

Aims of the Investigation

There are four factors involved in these tests, namely oils, engine makes, individual vehicles within makes, and months; and before an analysis is carried out it is necessary to decide what questions need answering in respect of these factors. The simplest question to answer would be: 'Was the average observed fuel consumption of the six vehicles on low viscosity oil higher or lower than that of the six vehicles on the standard grade during the twelve months of the tests?' The answer to this question would be obtained simply by averaging the fuel consumption corresponding to each lubricating oil and comparing the two, no test of significance being necessary. This question refers entirely to known facts, no generalisation of any sort being involved. A more

general question would be: 'Would the six vehicles on the low viscosity oil consume more or less fuel than would those on the standard grade when run over a long period of time?' Here a generalisation is involved from the specific twelve months during which the observations were taken to time in general. The most general question would be: 'Do vehicles when running on a low viscosity oil consume more or less fuel than when running on a high viscosity oil?' This involves prediction on all four factors, i.e. from individual vehicles to all vehicles, from two makes of engine to all such makes, from twelve months to time in general, and from two oils of low and high viscosity to oils of those viscosity classifications in general.

The three questions posed in the preceding paragraph will, it is hoped, have already drawn the reader's attention to the alternative interpretations that may be given to the levels of a factor in an analysis of variance. In one case the experimentalist is interested only in those specific levels at which he carries out his investigations and this type of factor is designated the 'specific' type in the discussion that follows. On the other hand he may be interested in a range of levels and those he actually investigates are simply a sample drawn from that range; this type of factor will be called the 'sampled' type. In the 'sampled' type the results are used to provide an estimate of the variance of the population from which the sample was drawn, whereas in the 'specific' type, since there is no implied population, there can be no variance and the observations provide estimates of a fixed series of constants associated with those specific levels. The way in which this distinction affects the tests of significance will be discussed later in the light of the actual analysis of the data under consideration.

Returning from the general to the particular, the question to be answered by this particular investigation was: 'Would a fleet of double-decker buses, consisting only of vehicles with engines A and B and at present running on the standard grade lubricating oil, benefit in fuel consumption if the oil were changed to the low viscosity oil being tested?'

It will be seen, therefore, that there is no difficulty in classifying three of the factors; *oils* and *engine makes* being of the 'specific' type and *individual vehicles* being of the 'sampled' type. The factor *months* on the other hand involves a more difficult decision; they are certainly not a sample of months drawn from the twelve months of the year, but on the other hand, they can be regarded as a sample of twelve consecutive months drawn from all possible months. In actual fact they are probably a mixture of the two types, the observed difference between January and February, say, being due to a fixed difference which will be repeated from year to year, together with a random variation which may differ from year to year.

Analysis of the Winter Data

Experience of similar work in the field of fuel consumption suggests

that the distribution of consumption rates for different vehicles of the same type is log-normal rather than normal, so that it is preferable to carry out the analysis of variance in terms of the logarithms of the consumption rates. This, of course, results in the tests of significance being based on percentage changes in consumption rates but this form of presentation of results is usually quite acceptable to all concerned.

Before the analysis of the winter data can be carried out, a decision has to be taken as to how to deal with the three missing values. The estimation of a missing value by minimisation of the error followed by a normal analysis of variance is, strictly, not valid; but in the case of one such value the departure from validity is generally accepted as being negligible. It is our opinion that the estimation of a number of missing values up to five per cent and possibly even ten per cent of the total number of results is justified on the score that the departure from strict validity of the conclusions arrived at in this way is outweighed by the resulting ease of analysis.

Reference to Tables I and II will show that the error in this case consists of the term *vehicles $\times$ months within oils and engines* and is the sum of the *vehicles $\times$ months* terms in each of the oils and engines groups. The minimisation of the error is therefore reduced to the minimisation of the *vehicles $\times$ months* term in the particular oil and engine group in which the missing value occurs. Fortunately each of the three missing values occurs in a different group and this fact still further simplifies the procedure. Considering, for instance, the ordinary grade/engine group; a consumption rate has to be estimated for vehicle 12 in November, the estimate being such as to minimise the interaction *vehicles $\times$ months* in this group. The interaction is a measure of the way in which the difference between vehicles 11 and 12 varies from month to month and it is clear that this variation will be minimised if the difference in November is made the same as the mean difference in the other five months. The data is to be analysed in logarithms and so the estimation has also to be carried out in logarithms; the process is illustrated in Table II.

TABLE II

Estimation of log consumption rate for vehicle 12 in November

Month	Log Consumption Rate		Difference	Estimation of x
	Vehicle 11	Vehicle 12		
November ..	2.049	x	$2.049 - x$	Mean difference $= -0.004$ $= 2.049 - x$ $x = 2.053$ (113 gallons/1 000 miles)
December ..	2.064	2.068	-0.004	
January ..	2.061	2.076	-0.015	
February ..	2.041	2.057	-0.016	
March ..	2.064	2.033	$+0.031$	
April ..	2.049	2.064	-0.015	

The estimate for vehicle 6 in November is obtained in exactly the same way but the estimation for vehicle 7 in April is slightly complicated

y the fact that there are four vehicles involved in this group instead of two. If, however, averages are used for the three vehicles whose data are complete, the process becomes the same as that in Table II, the log consumption rates for vehicle 11 being replaced by the average log consumption rates for vehicles 8, 9, and 10, and those for vehicle 12 being replaced by the log consumption rates for vehicle 7. Readers who are familiar with the concept of partitioning sums of squares into individual degrees of freedom will recognise this as the minimisation of the component (vehicle 7 versus remaining vehicles) $\times$ April versus other months), and a generalisation to higher order interactions is given in the appendix to this paper.

The winter data are now complete and the usual text-book analysis of variance can be carried out, bearing in mind that: (i) Three degrees of freedom are lost from the error term to compensate for the estimation of the three missing values, and (ii) The divisors for the crude sums of squares have to be adjusted to allow for the different numbers of vehicles in the sub-groups. The results are in Table III.

TABLE III

Analysis of variance of log consumption rate, winter data

Effect	$10^6 \times$ Sum of Squares	D.F.	$10^6 \times$ Mean Square	Expectation of Mean Square
Months (M) ..	2 890	5	578	
Engines (E) ..	11 700	1	11 700	
Oils (O) ..	21 115	1	21 115	$\sigma_0^2 + 6\sigma_v^2 + 3\cdot33\sigma_{moe}^2 + 6\sigma_{mo}^2 + 36\mathcal{Z}O^2$
ME ..	406	5	81	
MO ..	1 299	5	260*	$\sigma_0^2 + 3\cdot33\sigma_{moe}^2 + 6\sigma_{mo}^2$
OE ..	380	1	380	$\sigma_0^2 + 6\sigma_v^2 + 2\cdot67\sigma_{moe}^2 + 16\mathcal{Z}OE^2$
MOE ..	274	5	55	$\sigma_0^2 + 2\cdot67\sigma_{moe}^2$
Vehicles within E and O (Voe)	4 481	8	560†	$\sigma_0^2 + 6\sigma_v^2$
Error ..	3 687	37	100	σ_0^2

D.F. = Degrees of freedom. * Significant above the 5 per cent level. † Significant above the 0.1 per cent level.

The importance of deciding on the exact aims of the investigation now becomes apparent since the form of the last column of Table III, and hence the tests of significance, depend on this decision. It will be recalled that the two factors *engines* and *oils* were to be considered to be of the 'specific' type and so give rise to a series of constants rather than a component of variance. The distinction is drawn in Table III by the use of the notation $\mathcal{Z}O^2$ for the sum of squares of such constants for oils and $\mathcal{Z}OE^2$ for the interaction, as opposed to the usual σ^2 notation for components of variance. The expectation of the mean square for the *engines* main effect would have a similar $\mathcal{Z}E^2$ term if it were shown in the table but since the investigation is not concerned with the absolute difference between engines, the terms have not been evaluated. It will

be noted also that the term ΣOE^2 does not appear in the expectation of the mean square for the oils main effect and neither would it appear for the engines main effect. Disregarding for the moment the rather peculiar-looking coefficients the rules to be followed for formulating these expectations may be stated as:

- (i) A factor of the 'specific' type or an interaction consisting entirely of such factors gives rise to a term of the type ΣX^2 and this term appears once, and once only, in the analysis, namely in the expectation of the mean square corresponding to that effect.
- (ii) A factor of the 'sampled' type or an interaction consisting *entirely or partly* of such factors gives rise to a term of the type σ_x^2 and this term appears in the expectation of the mean square corresponding to that main effect or interaction and in every higher order interaction containing that effect.

The non-integer coefficients of the components of the expectations in this case are due to the fact that the analysis is 'orthogonal by proportionality.' It is generally assumed that for the data of a factorial design to be easily analysed, the number of observations for each level of any factor should be constant. This is to a certain extent true, since if the numbers of results vary haphazardly the usual techniques certainly cannot be applied; but if the numbers vary in such a way that, for any factor, the number of results at any one level is proportional to the number at any other level, then the analysis is said to be orthogonal by proportionality. Thus in the analysis of the winter data there are twice as many vehicles of engine-make A as of engine-make B on each of the lubricating oils in each month. Although the numerical computation of the analysis of variance table is relatively easy, as has been pointed out by Fairfield Smith,¹ proportionality leads to some complications in the tests of significance. Thus it will be seen from the analysis of Table III that if σ_{moe}^2 is found to exist and is believed to be an additional component of random error, then there is no direct F -test for the existence of σ_{mo}^2 because the coefficients of σ_{moe}^2 in the MOE and MO terms are not identical. Fortunately, this situation does not arise here for the MOE term is clearly non-significant when tested against the error term.

As would be expected, however, there are differences between the average consumption rates of the different vehicles, Voe being significant above the 0.1 per cent level, so that the existence of the component of variance σ_v^2 has to be accepted. The OE term may now be tested against Voe, and is clearly non-significant; so there is no reason to believe that the percentage change in consumption rate on make A differs from that on make B.

The MO term is just significant at the 5 per cent level when tested against the error, and it becomes necessary to decide whether, if σ_m is taken to exist, it represents a 'random' or a 'specific' component. For if it represents a 'specific' effect it should be struck out of the component of the O term which could then be tested against Voe directly.

procedure which will give a rough indication on this point is to break the OM term into the conventional polynomial time trend terms, as in Table IV.

TABLE IV
Breakdown of OM term, winter data

Effect		D.F.	Mean Square
O $\times$ M	Linear ..	1	51
	Quadratic ..	1	409
	Cubic ..	1	141
	Quartic ..	1	90
	Quintic ..	1	608
Error		37	100

There is no strong evidence of any physically likely trend here, so it is better to play safe and assume that the OM term represents an additional component of error. Under these circumstances there is no exact F -test for the O main effect, and resort must be made to the approximate procedures discussed by Cochran,² Johnson and Keeping,³ and others.

In a normal F -test the expectations of the numerator and denominator differ only by a multiple of the term whose existence is being tested. To test for the existence of the *oils* main effect therefore two terms are required whose expectations differ only by some multiple of EO^2 . Consideration of Table III will show that these may be obtained in a number of ways, of which two are illustrated in Table V.

TABLE V

Possible linear functions of mean squares for testing the existence of the oils main effect

	Numerator of F Ratio	Denominator of F Ratio
(a)	m.s. for O + m.s. for error with expectation: $2\sigma_0^2 + 6\sigma_v^2 + 6\sigma_{mo}^2 + 36EO^2$	m.s. for MO + m.s. for Voe with expectation: $2\sigma_0^2 + 6\sigma_v^2 + 6\sigma_{mo}^2$
(b)	m.s. for O with expectation: $\sigma_0^2 + 6\sigma_v^2 + 6\sigma_{mo}^2 + 36EO^2$	m.s. for MO + m.s. for Voe - m.s. for error with expectation: $\sigma_0^2 + 6\sigma_v^2 + 6\sigma_{mo}^2$

The question of which of these two tests should be applied is still a matter for discussion; Cochran himself suggests that (a) should be used, chiefly because it involves no negative terms; but we are of the opinion that in this particular case (b) is the more satisfactory for two reasons. (i) The number of degrees of freedom for one of the mean squares is already known and need not be found approximately by Cochran's method. (ii) One of the aims of the investigation is to place confidence limits to the change of fuel consumption and this is rendered fairly straightforward by the use of an F -test whose numerator has only one degree of freedom as is the case for (b).

Pooling the terms ME and MOE with the error to give a revised estimate of σ_0^2 of 93 with 47 degrees of freedom gives an F ratio of

$$F = \frac{21.115}{260 + 560 - 93} = 29.0$$

and the degrees of freedom of the denominator are calculated as

$$\phi = \frac{(260 + 560 - 93)^2}{260^2/5 + 560^2/8 + 93^2/47} = 10$$

giving a significant F ratio.

The conclusion to be drawn from these data is, therefore, that in winter the low viscosity lubricating oil leads to a reduction in fuel consumption and that there is no evidence that this reduction differs for the two engine-makes. The only remaining information required is an estimate of this reduction together with confidence limits. The estimation is perfectly straightforward and merely consists of taking the antilogarithm of the difference between the two average logarithms of consumption. This difference is -0.03425 , the antilogarithm of which is 0.924 giving the percentage reduction in consumption as 7.6 . Note that the difference must be taken in the order *low viscosity* minus *standard grade* since the change is to be expressed as a percentage of the consumption on the standard grade.

To obtain confidence limits for this reduction in consumption an estimate is required of its standard error, which will be some function of the three components σ_0^2 , σ_v^2 , and σ_{mo}^2 . Now the consumption figure corresponding to one of the lubricating oils was obtained by averaging 36 observations and hence σ_0^2 is divided by 36. Six vehicles were involved and so σ_v^2 is divided by 6, and there were six months so that σ_{mo}^2 is also divided by 6. The variance of an estimated consumption rate on one of the oils is therefore

$$\sigma_0^2/36 + \sigma_v^2/6 + \sigma_{mo}^2/6$$

and hence that of the difference between two consumption rates is

$$\sigma_0^2/18 + \sigma_v^2/3 + \sigma_{mo}^2/3$$

Reference to Table V will show that an estimate of 18 times the variance is already available as the denominator of the F -test for the oils main effect. The variance of the estimate of reduction in consumption is therefore $727 \times 10^{-6}/18 = 40.39 \times 10^{-6}$ with 10 degrees of freedom, corresponding to a standard error of 6.35×10^{-3} .

The 5 per cent point of t with 10 degrees of freedom is 2.23 leading to confidence intervals in logarithms of -0.03425 ± 0.01416 . On taking antilogarithms it is found that the improvement in fuel consumption in winter is 7.6 per cent with 95 per cent confidence limits of 4.5 and 10.6 per cent.

Analysis of the Summer Data

Referring back to Table I, it will be seen that the summer data are very much less complete than those for the winter months. In fact three vehicles are missing completely, and the most economical way of arriving at a proportional set of data is to omit vehicles 1, 6, and 8 as well. This still leaves two missing values in the set, one on vehicle 2 and the other on vehicle 5. The former can be estimated within the make A/low viscosity group by the method previously discussed, but the latter occurs in a group which now contains only one vehicle and this raises the question of what to minimise when the minimisation of the error cannot provide the estimate. The answer is that the estimate is obtained by minimising the highest order interaction that will provide that estimate, and the simplest way of deciding which interaction that is, is to carry out the following process on the data.

Sum the observed data as far as possible to give all the tables of sums necessary for the calculation of the analysis of variance. If any one such table has one or more cells completely blank, i.e. into which none of the observed data is to be put, then the sum of squares corresponding to that table will have to be minimised, subject to the condition that if that table contains any table corresponding to a lower order effect which also has a blank cell the sum of squares corresponding to that table must be minimised first. In the present case the missing value on vehicle 2 leaves a blank cell in the table corresponding to the *error* sum of squares (i.e. the original table) but leaves no blank cell in any subsidiary table and hence is estimated by minimising the error. The missing value on vehicle 5, on the other hand, although leaving a blank cell in the *error* table also leaves a blank cell in the MOE table and hence is estimated by minimising the sum of squares for MOE. The method of minimising a second order interaction is given in the appendix and will not be further discussed here.

The estimation of these two missing values entails the loss of two

TABLE VI

Analysis of variance of log consumption rate, summer data

Effect	$10^6 \times$ Sum of Squares	D.F.	$10^6 \times$ Mean Square	Expectation of Mean Square
Months (M) ..	1 151	5		
Engines (E) ..	2 048	1		
Oils (O) ..	4 727	1	4 727	$\sigma_o^2 + 6\sigma_v^2 + 1.67\sigma_{moe}^2 + 3\sigma_{mo}^2 + 18\sigma_o^2$
ME ..	1 890	5	378	
MO ..	2 666	5	533	$\sigma_o^2 + 1.67\sigma_{moe}^2 + 3\sigma_{mo}^2$
OE ..	612	1	612	$\sigma_o^2 + 6\sigma_v^2 + 1.33\sigma_{moe}^2 + 8\sigma_{oe}^2$
MOE ..	508	4	127	$\sigma_o^2 + 1.33\sigma_{moe}^2$
Vehicles within E and O (Voc)	67	2	34	$\sigma_o^2 + 6\sigma_v^2$
Error ..	710	9	79	σ_o^2

degrees of freedom, one from the error and one from the interaction MOE and the resulting analysis of the logarithms of the consumption rates is given in Table VI.

Surprisingly, in view of the winter results, the σ_v^2 term is not nearly significant for the summer data; and the usual procedure would lead to the conclusion that σ_{moe}^2 was non-significant, but that the ME, MO, and OE terms were all significant. From a physical viewpoint it appears unreasonable that in winter there should be differences between nominally identical vehicles operating under fixed service conditions, but that in summer these differences should vanish; and in view of the very limited (2 degrees of freedom) amount of information on σ_v^2 for the summer data it appears preferable to note that the OE term for summer would not be significant against the Voe term obtained from the winter analysis, and hence to regard the existence of ΣOE^2 representing a different rate of improvement for the two engine makes as 'not proven' for the summer data either.

The breakdown of the MO term again gives no clear evidence of systematic trends in the difference between the two lubricating oils, so that MO may again be regarded as an additional component of error. On incorporating the winter estimate of $\sigma_o^2 + 6\sigma_v^2$ (the two estimates of σ_o^2 are not demonstrably different), the F ratio for testing the average difference between the two oils becomes

$$F = \frac{4727}{533 + 560 - 94} = 4.73$$

the estimate 94 of σ_o^2 being obtained by pooling winter and summer estimates of σ_o^2 including the MOE terms which are non-significant in both cases. The degrees of freedom of the denominator in this case are estimated by

$$\phi = \frac{(533 + 560 - 94)^2}{533^2/5 + 560^2/8 + 94^2/55} = 10.3$$

The observed mean difference in the logarithms of the rate of consumption for the summer data is -0.02456 , and the 95 per cent confidence limits obtained as before are -0.02456 ± 0.02525 . On taking anti-logarithms, the estimated improvement in consumption for the summer months becomes 5.5 per cent, with 95 per cent confidence limits of -0.2 and 10.8 per cent.

Joint Analysis of Summer and Winter Data

Having decided, albeit somewhat uneasily in view of the discrepancies between them, on the conclusions to be drawn from the separate analyses, one major point at issue is still outstanding. Is the gain in consumption in winter different from that in summer? This point would in theory be quite easily answered by a joint analysis if the data were complete, but the scanty nature of the summer data upsets matters

considerably. If these were appreciably less scanty, a joint analysis in which all the gaps were filled by missing value technique would be quite feasible, but as previously indicated we draw the line at estimating more than ten per cent of the total data and prefer to estimate not more than five per cent.

Again, had the winter vehicles been entirely distinct from the summer vehicles, a joint analysis of the data so far examined would have been straightforward and the results would have been orthogonal by proportionality, there being twice as many vehicles of make A as of make B and twice as many of each in winter as in summer. The fact that some of the vehicles are common to both sets of data, however, invalidates this approach and means that the winter and summer estimates of gain in consumption are non-independent. The best simple approach is to determine the covariance between these two estimates and then derive the variance of their difference and this forms the next part of the investigation.

The covariance arises out of the fact that some of the vehicles are common to both sets of data and must therefore consist only of the term σ_v^2 . Considering only the *vehicle* terms in the comparison, the gain in consumption in winter may be represented as

$$(v_1 + v_2 + v_3 + v_4 + v_5 + v_6)/6 - (v_7 + v_8 + v_9 + v_{10} + v_{11} + v_{12})/6$$

and that in summer by

$$(v_2 + v_3 + v_5)/3 - (v_7 + v_9 + v_{11})/3$$

being used for vehicle so as to avoid confusion with the later notation for variance. The product of these two expressions is

$$(v_2^2 + v_3^2 + v_5^2 + v_7^2 + v_9^2 + v_{11}^2)/18 + \text{the sum of products of unlike } v\text{'s}$$

Since the vehicles are independent the sum of products of unlike v 's provides no contribution to the covariance and this is therefore estimated as

$$C_{ws} = (6\sigma_v^2)/18 = \sigma_v^2/3$$

The variance of the winter estimate has already been derived as

$$V_w = \sigma_o^2/18 + \sigma_v^2/3 + \sigma_{mo}^2/3$$

and that for the summer estimate may be similarly derived as

$$V_s = \sigma_o^2/9 + 2\sigma_v^2/3 + \sigma_{mo}^2/3$$

Adding to a variance of the difference between the two estimates of

$$V_w + V_s - 2C_{ws} = \sigma_o^2/6 + 2\sigma_{mo}^2/3 + \sigma_v^2/3$$

which may be represented as a linear function of σ_o^2 , $\sigma_o^2 + 6\sigma_v^2$, and σ_{mo}^2 .

Unique estimates of σ_o^2 and $\sigma_o^2 + 6\sigma_v^2$ are available by pooling these for winter and summer, but there are two estimates of σ_{mo}^2

through the $\sigma_0^2 + 6\sigma_{mo}^2$ estimate for winter and $\sigma_0^2 + 3\sigma_{mo}^2$ for summer and therefore an infinite number of estimates of the above function from which some 'best' estimate has to be chosen.

If we accept as best the estimate that is subject to the least variance, the one required is that with the greatest number of degrees of freedom. As an approximation, therefore, the formula given by Cochran for the degrees of freedom of a linear function of a number of estimates may be maximised with respect to the coefficients, subject to the condition that these are such that the linear function estimates the required variance. This, as shown in the Appendix, leads to a linear equation in one of the coefficients and the best estimate of the variance of the difference between the two observed gains in consumption is found to be 73.53×10^{-6} (in terms of logarithms) with 17 degrees of freedom. The difference between the logarithms of the percentage increases is 0.009 69 which, with a standard error of 0.008 6, is not significant. There is therefore no evidence that the gain in consumption in winter is different from that in summer.

Having decided this, the summer and winter estimates may be pooled to obtain the most efficient estimate of the increase in consumption. At first sight, there are two estimates of the same quantity each with its own variance, and pooling these to give one estimate might appear to be perfectly straightforward. If the two estimates were independent, the most efficient joint estimate of the gain in consumption would be the weighted mean of the summer and winter estimates, the weights being inversely proportional to the variances. If only estimates of these variances are available then recourse can be made to Meier,⁴ who deals with such a situation.

In the present case, however, the two estimates of the gain are not independent but have a covariance of $\sigma_v^2/3$ and the problem arises of finding the most efficient joint estimate from two non-independent estimates. As an approximation this may be done (see the Appendix) by considering the estimates for the summer and winter gains to be independent with variances of $V_s - C_{ws}$ and $V_w - C_{ws}$ respectively, the notation being that used above, and this leads to the following variances for the two estimates:

$$\text{for summer } V_s - C_{ws} = \sigma_0^2/18 + \sigma_{mo}^2/3$$

$$\text{for winter } V_w - C_{ws} = \sigma_0^2/9 + \sigma_{mo}^2/3 + \sigma_v^2/3$$

Again, as in estimating the variance of the difference between the two gains in consumption, two estimates are available for σ_{mo}^2 and the best estimates are required of the two variances. Proceeding as before and maximising Cochran's formula for the degrees of freedom leads to the following numerical values:

$$\text{estimate of } V_s - C_{ws} = 10^{-6} \times 54.84 \text{ with 16 degrees of freedom}$$

$$\text{estimate of } V_w - C_{ws} = 10^{-6} \times 21.64 \text{ with 9 degrees of freedom}$$

and applying Meier's technique gives a joint estimate of the logarithm of the gain in consumption as -0.03151 with a variance of 17.70×10^{-6} with 16 degrees of freedom. On taking antilogarithms the final estimate of the gain in consumption on changing from standard grade to low viscosity oil is found to be 7.0 per cent with 95 per cent confidence limits of 5.0 and 8.8 per cent.

Conclusion

Some further analyses were carried out on the data to investigate the possible existence of cyclical trends in fuel consumption but these, though interesting theoretically, were considered to be outside the scope of this paper. The work already discussed provides enough examples of the ways in which a very ordinary set of data can result in deviations from the usual text-book types of analysis. The industrial statistician may seldom be in a position to devote as much time to any one job as was involved in carrying out these analyses, but on the other hand it is probable that the majority of problems which he tackles will in fact be subject to one or more of the common causes of invalidity and the question arises of how far it is necessary and desirable to attempt to eliminate such causes by the adoption of more complex and time-consuming methods of analysis.

A considerable amount of work, both theoretical and empirical, has been carried out to determine the effects of non-satisfaction of the assumptions underlying the analysis of variance, but it will be noted that this work is directed mainly at the validity of the tests of significance, and in fact is directed largely at determining the effects of non-normality. We believe that the information so derived is of secondary importance to the practising statistician whose primary concern lies in the estimation of the parameters of the model and of their standard errors.

If a complex method of analysis is believed to be necessary to extract all the information from a set of data, the question of whether it should be adopted will clearly depend on the urgency and importance of the problem. In our view the best procedure is to make the most of conventional analyses so as to give a reserved judgement in the first place. If the problem justifies more complex analysis it will often be worthwhile to collect additional data in a form which will best suit the type of analysis envisaged. Inevitably, the decision to proceed must be based largely on the hunch and experience of the statistician concerned, but further theoretical work on such situations would be of considerable value in taking these decisions out of the 'shots in the dark' class.

Acknowledgement

We are indebted to those colleagues who were responsible for the conduct of this investigation, and to the Shell Petroleum Company Limited for permission to publish this paper.

APPENDIX

Estimation of Missing Value

The logarithms of the summer consumption data as used in the analysis are given in Table 1, the missing value being denoted by x .

Table 1
 $10^3 \times (\text{Logarithms of summer consumption data} - 2)$

Oil	Low Viscosity			Ordinary Grade		
Engine Make	A		B	A		B
Vehicle No.	2	3	5	7	9	11
May	57	45	x	53	57	49
June	33	37	29	68	93	45
July	41	49	25	65	68	49
August ..	41	41	64	68	57	41
September ..	53	57	45	90	79	61
October ..	41	61	9	100	104	57

The interaction $O \times E$ has to be minimised with respect to x , and this interaction can be partitioned into individual degrees of freedom such that x appears in only one of these, the interaction will be minimised by reducing this particular component to zero. The only such system of partitioning is

$O \times E \times (\text{May versus other months})$ with 1 degree of freedom

$O \times E \times (\text{Comparisons within other months})$ with 5 degrees of freedom

The first component can be written in a two-way table as in Table 2 where the data for engine make A are the means for vehicles 2 and 3 and 7 and 9 respectively.

Table 2
Interaction $O \times E \times (\text{May versus other months})$

(5 $\times$ difference between May data and mean of other monthly data)

Engine Make	A	B	Difference
Low viscosity oil ..	28	$5x - 172$	$5x - 200$
Ordinary grade oil ..	- 121	- 8	113

The component $O \times E \times (\text{May versus other months})$ is reduced to zero by solving the equation $5x - 200 = 113$ giving $x = 63$. Thus log consumption for vehicle 5 in May is 2.063 which is equivalent to 115 gallons per 1000 miles.

The process is illustrated on a second order interaction, two of whose factors are at two levels, but the procedure is very easily extended to

second order interactions in general. An interaction ABC, say, where A is at levels $A_1 A_2 \dots A_n$, B at levels $B_1 B_2 \dots B_m$, C at $C_1 C_2 \dots C_p$, may be minimised with respect to a missing value at $A_1 B_1 C_1$, by reducing to zero the component

$$(A_1 \text{ versus other A's}) \times (B_1 \text{ versus other B's}) \times (C_1 \text{ versus other C's})$$

Briefly, this is done by representing this interaction in a two-way table (A $\times$ B say) of differences between the values at C_1 and the mean of the values at the other C's and then proceeding as for first order interactions as shown earlier in this paper.

The generalisation to still higher order interactions is obvious.

Maximisation of Degrees of Freedom

As shown in the paper the variance of the estimate of the difference between the summer and winter gains in consumption is $(\sigma_0^2 + 4\sigma_{mo}^2 + 2\sigma_v^2)/6$ and there are available four estimates of components of variance which are to be combined in some linear function to provide an overall estimate of this variance. These four estimates are given in Table 3.

Table 3
 $10^6 \times$ Estimates of components of variance

Component	Estimate	Degrees of Freedom
σ_0^2	$S_0^2 = 94$	$f_0 = 55$
$\sigma_0^2 + 6\sigma_v^2$	$S_1^2 = 560$	$f_1 = 8$
$\sigma_0^2 + 3\sigma_{mo}^2$	$S_2^2 = 533$	$f_2 = 5$
$\sigma_0^2 + 6\sigma_{mo}^2$	$S_3^2 = 260$	$f_3 = 5$

The linear function

$$S^2 = aS_0^2 + bS_1^2 + cS_2^2 + dS_3^2$$

will provide the required estimate provided

$$\sigma_0^2 + b(\sigma_0^2 + 6\sigma_v^2) + c(\sigma_0^2 + 3\sigma_{mo}^2) + d(\sigma_0^2 + 6\sigma_{mo}^2) = (\sigma_0^2 + 4\sigma_{mo}^2 + 2\sigma_v^2)/6$$

i.e. provided $b = 1/18, c = -2a, d = a + 1/9$, and hence

$$S^2 = aS_0^2 + S_1^2/18 - 2aS_2^2 + (a + 1/9)S_3^2$$

As an approximation to the degrees of freedom of this estimate, Cochran's formula gives

$$\phi = \frac{[aS_0^2 + S_1^2/18 - 2aS_2^2 + (a + 1/9)S_3^2]^2}{a^2S_0^4/f_0 + S_1^4/(18^2f_1) + 4a^2S_2^4/f_2 + (a + 1/9)^2S_3^4/f_3}$$

and a may be so chosen as to maximise ϕ .

Now ϕ is of the form

$$\phi = \frac{(m + na)^2}{p + 2qa + ra^2}$$

and this may easily be shown to have a maximum value of

$$\phi_{\max} = \frac{2mnq - rm^2 - pn^2}{q^2 - rp} \text{ at } a = \frac{mq - np}{nq - rm}$$

Substitution of the numerical values from Table 3 gives

$$\phi_{\max} = 17.0 \text{ at } a = -0.0190$$

and

$$S^2 = 73.53$$

A Best Joint Estimate from Two Correlated Estimates

Two unbiased but correlated estimates x_1 and x_2 are available of an unknown quantity θ which are to be combined in some linear form to provide the best single estimate of θ .

Consider the linear function

$$X = a_1x_1 + a_2x_2$$

where, since X is to be an unbiased estimate of θ , $a_1 + a_2 = 1$. The variance of X is

$$\text{Var}(X) = a_1^2V_1 + a_2^2V_2 + 2a_1a_2C_{12}$$

where V_1 and V_2 are the variances of x_1 and x_2 respectively and C_{12} their covariance. Minimising this with respect to a_1 and a_2 subject to the condition that $a_1 + a_2 = 1$ gives

$$a_1 = \frac{V_2 - C_{12}}{V_1 + V_2 - 2C_{12}}, \quad a_2 = \frac{V_1 - C_{12}}{V_1 + V_2 - 2C_{12}}$$

and hence

$$\begin{aligned} X &= \frac{(V_2 - C_{12})x_1 + (V_1 - C_{12})x_2}{V_1 + V_2 - 2C_{12}} \\ &= \frac{x_1/(V_1 - C_{12}) + x_2/(V_2 - C_{12})}{1/(V_1 - C_{12}) + 1/(V_2 - C_{12})} \end{aligned}$$

The best single estimate therefore is obtained by weighting the two estimates x_1 and x_2 by the inverses of the quantities $V_1 - C_{12}$ and $V_2 - C_{12}$ respectively. Now the best linear estimate obtained from two *independent* estimates is that obtained by weighting inversely as the variances, and Meier has evolved a method for using estimates of these variances when their actual values are unknown. If, therefore, the two correlated estimates are treated as independent with equivalent variances of $V_1 - C_{12}$ and $V_2 - C_{12}$ respectively, Meier's method can be applied.

REFERENCES

- FAIRFIELD SMITH, H. (1951). 'Analysis of variance with unequal but proportionate numbers of observations in the sub-classes of a two-way classification', *Biometrics*, **7**, 70.
- COCHRAN, W. G. (1951). 'Testing a linear relation among variances', *Ibid.*, **7**, 17.
- JOHNSON, L. P. V. and KEEPING, E. S. (1952). 'Composite mean squares and their degrees of freedom', *Applied Statistics*, **1**, 202.
- MEIER, P. (1953). 'Variance of a weighted mean', *Biometrics*, **9**, 59.

LETTER TO THE EDITOR

A Point in the Construction of Price Index Numbers

HAVING read Dr Ruderman's article 'A neglected Point in the Construction of Price Index Numbers' in this journal (3, 1954, 44) at its proof stage, and having had no previous opportunity of suggesting any modification, I feel obliged to clarify my own viewpoint by demonstrating it in more general terms.

Let us, according to Dr Ruderman's notation with slight modification, use the subscript s to identify store prices of the index items, and f for all the varieties of items of a certain pattern of family expenditure. P_s refers to a certain specification of an item of commodities or services, whereas P_f is a 'hypothetical' average price of the item, since in practice $P_f Q_f$ cannot be factored. Q_s is also a 'hypothetical' quantity defined as $Q_{s0} = P_{f0} Q_{f0} / P_s$ or $Q_{f0} = Q_{s0} / A_0$, where P_{f0} differs from P_{s0} by the ratio A_0 (not exactly corresponding to Dr Ruderman's a). Supposing $A_0 = A_t$ when $P_{f0} = A_0 P_{ft}$ and $P_{ft} = A_t P_{st}$. The relative formula, expressed as $\Sigma(P_{st}/P_{s0})P_{f0}Q_{f0}/\Sigma P_{f0}Q_{f0}$, can be transformed into $\Sigma(P_{ft}/P_{f0})P_{f0}Q_{f0}/\Sigma P_{f0}Q_{f0}$ and, consequently, into $\Sigma P_{ft}Q_{f0}/\Sigma P_{f0}Q_{f0}$, where P_{f0} , P_{ft} , and Q_{f0} refer to all varieties of commodities and services in a specific consumption pattern. On the other hand, the aggregative formula in common practice, being expressed as $\Sigma P_{st}Q_{s0}/\Sigma P_{s0}Q_{s0}$, is also based on the same assumption, viz. $A_0 = A_t$, so that the aggregative formula can be a real substitute for the ultimate formula $\Sigma P_{ft}Q_{f0}/\Sigma P_{f0}Q_{f0}$.

Q_s represents a hypothetical quantity, and $P_f Q_f$ cannot be factored in practice. These statistical values being of such a nature, the indifference curve approach as a theoretical basis of the index formula cannot but be limited even apart from its unreal assumption of unchanged consumers' preferences. In other words, $P_{s0}Q_{s0}$ or $P_{st}Q_{s0}$, which appears in the aggregative expression, does not represent an expenditure line associated with real combinations of consumption goods on the indifference map.

The weighted mean of price relatives is popularly used not only for practical reasons but also because of its own economic meaning, which is equally well founded as the aggregative expression on the theoretical basis.

ITASU SAKURAI

24th November 1954

[This correspondence is now closed.—Editor]

QUESTION AND ANSWER

Readers are invited to submit questions or to deal shortly with points of general interest by supplying both the question and the answer.

Question

My firm produces one article in some hundreds of types, so that any one type is produced only once every month or six weeks. What stock of any one type should be held if customers are to be kept waiting only once in twenty periods? Are there any references on the subject, and can a method be produced for which the working is sufficiently simple to supplement the present *ad hoc* method?

Answer

Quite a lot of attention has been given to this, the so-called 'inventory problem', but usually from the economic point of view. References to the key papers, the last of which gives a bibliography of the subject, are given below.^{1, 2}

In order to answer the question one must of course have knowledge of the distribution of the demand in any six-week period. This could, if thought worth while, be corrected for seasonal trends so that demand is expressed as the sum of a seasonal component D_i (constant for a given season i of the year) and a variable component d which includes all the remaining sources of variability. For a simple solution we need from this knowledge of the distribution to choose a value $d_{0.95}$ which is only exceeded in 5 per cent of periods and to make up the stock to $D_i + d_{0.95}$.

This oversimplifies the problem because usually there is constant capacity available which will be switched to whatever item is required as soon as the previous batch has been produced. Thus the intervals over which demand is required to be predicted are not equal. Nor is the cost of storage to be balanced against the cost of refusing demand.

If, however, one is prepared to simplify to this extent I would recommend preparing estimates of demand for periods of one month, this being probably long enough to cover preparation of order, planning, and completion. These estimates would vary according to season and to type and would be in the form of a table. Each week the stock would be compared with these estimates and orders placed for those types for which the stock was less than or close to this estimate. The size of the order would be that required to bring the stock up to the $D_i + d_{0.95}$ level corresponding to some longer period learnt by experience after the scheme had been in operation, and also available as a table. On the information given this period would be six weeks from the completion of the order, and the size of the order would also need to make some allowance for depletion of stock pending completion of order. Orders could be given a priority according to the size of the deficiency below

estimate and, if factory procedure allowed, these priorities could be altered during the planning stage as a result of the subsequent weeks' stock examination.

Note that the two estimates may have different seasonal components. The 'warning to order' estimate will refer to the month immediately to follow whereas the 'size of stock' estimate will be that for the ten weeks immediately following which should desirably elapse before another order is required. It is worth noting that the frequency distribution of d , the random component in the demand for a given period, might be estimated from the distribution of demand for shorter periods simply by the addition of variances.

¹ DVORETSKY, A., KIEFER, J., and WOLFOWITZ, J. (1952). 'The inventory problem' *Econometrica*, **20**, 187 and 450; see also **21**, 586.

² ARROW, K. J., HARRIS, T., and MARSCHAK, J. (1951). 'Optimal inventory policy' *Ibid.*, **19**, 250.

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

List and Reports of Meetings

BIRMINGHAM AND DISTRICT GROUP

8th October 1954. Productivity Studies in the Cotton Industry: L. H. C. Tippet

Mr Tippet described the various types of factory study carried out in the cotton industry and showed how an index of productivity was derived. He discussed the techniques employed in these studies, such as time study, snap readings, examination of factory records, etc. Finally he showed how the work was tied up with the fixing of a wage system and prescriptions for mill efficiency.

20th October 1954. Visit to the Guest, Keen and Nettlefolds Group Research Laboratories, Wolverhampton

Dr T. Emmerson, Director of Research, prefaced the tour of the new laboratories by describing the rapid post-war development of the research department and its place in the Group. He gave examples of some important operational research projects, showing how statistical techniques combined with scientific method had brought results that were valuable to factories in the Group.

27th November 1954. Operational Research during the War: P. M. S. Blackett

After giving his own definition of operational research, Professor Blackett traced its development by describing how scientists went into operational units of a number of branches in the armed services, where they became 'trouble-finders'. He illustrated how operational research helped to solve problems arising from the war against submarines, the utilisation of aircraft, the best size of convoys, and air-bombing.

GLASGOW GROUP

22nd October 1954. Developments in Operational Research. A General Survey: J. C. Eaton

11th November 1954. Ratio Delay Techniques: J. H. Hepburn

12th December 1954. The Use of Graphical Statistics in Steelmaking: D. Hadfield

LEICESTER GROUP

11th November 1954. The Development of a Statistical Investigation: D. J. Desmond

12th December 1954. Audience Research for the BBC: R. Silvey

LONDON GROUP

1st October 1954. Statistical Methods and Operational Research in the Motor Industry: A. Hardman

Mr Hardman gave a general review of the different ways in which a statistician can assist management in making decisions, and outlined a number of specific applications. Among the latter were the analysis of trade association statistics and the published accounts of rival undertakings in order to assess the comparative 'health' of one's competitors. The co-ordination of production planning with time study, machine loading, labour recruitment, and stock control was also discussed, and the speaker ended with a light-hearted assessment of probable collision rates from a study of the number of cars on the roads at any time.

29th October 1954. Statistics and the Accountant—2: A. Muir. Joint Meeting with the London Branch of the Institute of Cost and Work Accountants

After pointing out the scarcity of published material on the application of statistical methods to accounting, Mr Muir cited some cases where he had found them of value. The examples included the use of least squares in setting up models of when debtors settled their accounts, the comparison and analysis of production expenses by regression methods, the sampling of accounting data, and the control of costs.

MERSEYSIDE GROUP

5th October 1954. The Statistical Approach to Weights and Measures: D. J. Desmond

This paper was substantially the same as that given in Birmingham in April 1954 and reported in *Applied Statistics*, 3, 200.

2nd November 1954. The Estimation of Expected Life from Incomplete Forced Life-test Data: Miss J. Keen

Miss Keen discussed a problem arising in the testing of insulators. Under normal conditions these should last for fifty years, but under more severe conditions the average life could be reduced to eight years. A number of sets of insulators were run at varying voltage and temperature conditions and time to failure was measured. If it was assumed that the lives were distributed log-normally, means could be estimated for sets in which only two or three insulators had actually failed. A multiple regression equation was fitted relating $\log(\text{life})$ to temperature and voltage, and the optimum temperature and voltage condition selected for future testing.

7th December 1954. Non-parametric Tests: G. A. Barnard

This paper was substantially the same as that given to the South Wales Group in January 1954 and reported in *Applied Statistics*, 3, 132.

NORTH-EASTERN GROUP

13th October 1954. The Development of a Statistical Investigation: D. J. Desmond

Mr Desmond quoted two examples where quality control techniques were sufficient to improve quality and reduce wastage. He then described a mor

complex investigation in which errors in measurements were found to be of comparable magnitude to the specified tolerance. He pointed out the difficulties associated with a flow-line production and gave a method of establishing a control scheme. The practical consequences of the investigation were improved quality at a reduced cost and the ability to 'loosen' the specification.

11th November 1954. Figures and Facts: B. P. Dudding

Dr Dudding described the use of a simple range chart in showing the need for process correction. He then gave an example where a machine shop efficiency had been calculated on average values, which had hidden the need for adjustments on individual machines. Improved overall efficiency resulted from charting these average values, but it could have been even better if more attention had been paid to individual machines.

SHEFFIELD GROUP

14th September 1954. Some Examples of Operational Research in the Coal Industry: B. H. P. Rivett

Three problems were solved by cost studies: dirt disposal, a comparison of the efficiencies of diesel and battery locomotives, and the comparative costs of different forms of underground haulage. Because the data were collected from pits where the situations were not comparable, the costs were broken down by regression analysis and a total cost synthesised for a typical situation. In a fourth problem the telephone systems of several mines were studied and distributions of waiting times obtained by random numbers. A system of exchanges and calls was then devised which would ensure minimum waiting times for emergency calls.

17th October 1954. The Analysis and Use of Industrial Data: Miss J. Keen

Miss Keen described the use of inspection schemes in raw material departments and mentioned some of the objections raised by management. She discussed process control, the use of control charts, regression analysis and analysis of variance was described, and examples were given. Brief reference was made to queueing methods, Tippett's snap-reading method, and the application of regression analysis to sales forecasting.

SOUTH WALES GROUP

19th October 1954. Statistical Methods applied to a Study of Incentives: Mr L. T. Wilkins

22nd November 1954. Small Particle Statistics: G. Herdan

1st December 1954. A Practical Approach to making Statistical Decisions: Mr G. A. Barnard

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

Reports of Meetings

BRISTOL GROUP

21st October 1954. The Use and Presentation of Statistics: G. F. Todd

Mr Todd listed the particular requirements for the presentation of statistical information about an industrial or commercial concern, and pointed to the general principles governing the construction of statistical reports and the use of raw data to form them. He then illustrated the problems of the sound presentation of data having reference to the most important points they convey, and commented on the right use of charts and diagrams, etc.

18th November 1954. The Uses and Abuses of National Income Statistics: R. C. Tress

Professor Tress gave a comprehensive historical survey of the progress of national income accounting and the varying methods of presentation used, discussing all the relevant publications. This was followed by comments on the use and application of the data.

STUDY CIRCLE ON MEDICAL STATISTICS

28th September 1954. Chairman's Address: W. P. D. Logan

Dr Logan described the work of the General Register Office in the field of mortality and morbidity statistics. He outlined some of the department's plans for the immediate future and discussed, among other problems, the analysis of occupational and social mortality rates.

26th October 1954. Research in General Practice: J. Fry

A full account of this meeting, in which Dr Fry described his own experience in recording and analysing material collected in a general practice, is contained in the *British Medical Journal* (1954), 2, 1100.

30th November 1954. The Teaching of Medical Statistics: L. Bernstein

Dr Bernstein gave a fascinating account of the pioneer work at the London Hospital, where first-year medical students are being given practical instruction in the application of statistical methods to physiology and pharmacology. It is hoped that a paper on this subject will appear in a later issue of *Applied Statistics*.

BOOK REVIEWS

L. Davies (editor): *The Design and Analysis of Industrial Experiments*. Edinburgh: Oliver and Boyd, 1954. Pp. xiii + 637. 63s.

This book is certainly a valuable addition to statistical literature. The majority of books on experimental design have been written more especially for the agricultural worker, and there has been a definite need for a book addressed specifically to the industrial experimenter, written in terms that he can readily understand and appreciate, and including treatment of the problems that are peculiar to his field of activity. There are, indeed, one or two other books on industrial experimentation that have been published during the last few years: but there is nothing in the least comparable with the present work, which combines the qualities of being really sound, well written, comprehensive, and up to date. The book has been written principally for the research worker in the chemical industry; but it should also be useful to persons in related industries, and to lecturers and students in universities and technical colleges. The reader is assumed to have a limited knowledge of methods of statistical analysis: these are described in the authors' previous book, *Statistical Methods in Research and Production*, to which frequent references are made.

The subject matter has been set out on more or less conventional lines, starting with the presentation and discussion of fundamental concepts and basic methods concerning both design and analysis, and continuing systematically with the various types of design—randomised blocks and Latin squares, incomplete blocks, factorial designs, confounding, and fractional replication. The treatment throughout is clear and thorough, and every important point has been illustrated by a really good practical example. In particular, the authors are to be congratulated for having nicely judged the requisite degree of mathematical detail and basic theory.

The final chapter is devoted to the design of experiments for the determination of optimum conditions: as stated in the preface, 'Suitable designs have recently been introduced (by the ICI workers themselves, and mainly Dr Box), and a full treatment of the methods and examples of their application in practice are a special feature of this book'. This material is an outstanding contribution to the progress of industrial experimentation and should be studied carefully by every worker in the field of process development.

Mistakes and misprints are very few indeed and are relatively unimportant. On page 131, referring to the analysis of an unbalanced three-fold nested classification, the coefficient given for the variance component σ_1^2 in the seven batches expected mean square (Table 4A.2) is wrong; though fortunately this is not likely to lead to any serious errors in practice except in the lack of balance is very marked. The correct expression for this coefficient has been given by Ganguli (1941) in *Sankhya*, 5, 449–52.

D. R. READ

E. S. Pearson and H. O. Hartley (editors). *Biometrika Tables for Statisticians*, Volume I. Published for the Biometrika Trustees by the Cambridge University Press, 1933. Pp. xiv + 238. 25s.

Part I of the original Biometrika tables with the title *Tables for Statisticians and Biometricians* was first issued in 1914; four reissues have since appeared, the last in 1948—without alteration in its form. These tables were widely used by statisticians and were the only ones in existence in book form until Fisher and Yates issued their *Statistical Tables for Biological, Agricultural and Medical Research* in 1938. Owing to the considerable development of statistical methods during the past thirty years, many of the tables had become obsolete and new ones were required, but the fact that these tables had survived so long a life, much of it during this period of rapid expansion, is surely a high tribute to their value and to their editor, Karl Pearson.

Volume I of the new *Biometrika Tables* represents a complete recasting of its forerunner; twelve of the most commonly used tables have survived, but there are no fewer than forty-two new ones. The more common statistical functions, e.g. the Normal curve, the t -distribution, the χ^2 -distribution, the distribution of the range, the B-distribution, etc., are fully tabulated, giving not only the probability integral but also the percentage points of the distribution.

A valuable addition is a detailed table of probits. It is interesting to note that the χ^2 -table is expressed in such a form as to give directly the partial sums of the Poisson distribution. Among the further useful additions of special value to industrial statisticians are tables for comparing means of two groups whose variances are different and have to be separately estimated, the studentised range, charts and tables for the confidence limits of the correlation coefficient, confidence limits of the standard deviation, exact tests of significance in 2×2 frequency tables, confidence limits in sampling from binomial and Poisson distributions, tables of powers, factorials, etc.

A substantial introduction of ninety-seven pages precedes the tables. The various uses of each table are described in simple terms and amply illustrated by numerical examples. This is no ordinary introduction; indeed, under its guidance it can serve as an instruction to assistants on a wide range of statistical methods, and to a statistician it may serve as a useful refresher on many points.

The volume has catered primarily for the requirements of statisticians in the analysis of data, but the more fundamental functions have been tabulated to a greater accuracy than is normally required in order to satisfy the needs of those concerned with mathematical development. The emphasis is on the analysis of data, and apart from a most useful table of orthogonal polynomials, the needs of experimental design are not catered for, nor is the proposed for the second volume which is being compiled. This is not intended as a criticism but merely as a statement of fact. It is hoped that attention will be given to this particular need of industrial statisticians.

The standard of reproduction is excellent, as are also the quality of the paper and the binding—important matters in a book that will be in almost constant use—and the sizes of the print and the tables are nicely judged. This volume of the *Biometrika Tables* is clearly a 'must' for all statistical research units.

O. L. DAVENPORT

RESERVOIRS OF INFECTION OF PULMONARY TUBERCULOSIS IN WARWICKSHIRE

MARJORIE JESSOP

In giving results showing the incidence of tuberculosis in Warwickshire Mrs Jessop illustrates what useful information may be derived from a simple analysis of data in the possession of local authorities.

EXPERIENCE has shown that a statistician working in the Public Health Department of a local authority will only rarely have need to apply the more complicated statistical methods. However, the usefulness of the results which can be achieved by the most simple methods goes a long way towards compensating him for the lack of purely statistical interest. This note describes a statistical investigation made to discover which groups of the population in Warwickshire have a high incidence of pulmonary tuberculosis and constitute the most dangerous foci of infection. The methods used are extremely simple, consisting mainly of fitting up the data in various ways and calculating appropriate rates. With one important exception, all the information was obtained from routine returns made to the Health Department. The investigation is described in the hope that it will encourage more people to make use of the vast mass of routine returns accumulated in local government and other similar offices.

data

When a new case of pulmonary tuberculosis is diagnosed, a statutory form of notification is used to provide information for the public health authorities. This gives the name, age, sex, and place of residence of the person concerned. Notification of the disease when diagnosed is thought to be reasonably complete. There are, unfortunately, many persons who do not have the disease but have not yet been discovered and notified. It was to help in finding these cases that this investigation was made. One additional and very important piece of information has been collected for new cases diagnosed in 1952 and 1953. This is the stage of disease at diagnosis as assessed by the chest physician attending the case. The classification used is one devised by the Ministry of Health for use in returns from hospital authorities. The letters RA are used to denote cases in which no tubercle bacilli were found in the sputum (sputum negative cases) and RB to denote cases in which tubercle bacilli were found in the sputum (sputum positive cases). Sputum negative cases are regarded as non-infectious, and sputum positive cases as infectious. These two main groupings are further divided into three stages: 1, 2, and 3, according to the extent of damage to the lungs.

Stage 1 can be regarded as early, stage 2 as intermediate, and stage 3 as late.

Sex and Age Incidence

The numbers of new notifications in the five years 1949-53 were first divided into sex- and age-groups. It was found that there was little change from year to year in the numbers in each sex- and age-group so the results are presented here as totals over the five-year period. With the aid of the one per cent sample tables of the 1951 census, which give the sex- and age-distribution of the population of Warwickshire, the average yearly incidence of new notifications has been calculated for each age-group. It is shown in Fig. 1, with the numbers of notifications on which the diagram is based given beneath.

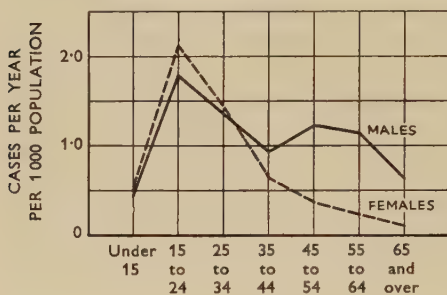


FIG. 1

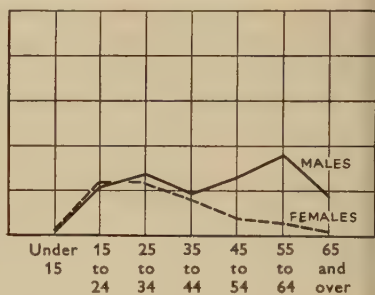


FIG. 2

FIG. 1. Age-incidence of new cases of pulmonary tuberculosis in Warwickshire, 1949-53. The numbers of notifications on which the points in the diagram are based are, in order:

Males:	148	260	242	200	196	122	64	total	1 232
Females:	158	307	255	130	63	33	15	total	961

FIG. 2. Age-incidence of new positive sputum cases of pulmonary tuberculosis in Warwickshire, 1952-53. The numbers of notifications on which the points in the diagram are based are, in order:

Males:	3	32	47	40	42	38	17	total	219
Females:	1	34	40	32	13	8	3	total	131

Of the approximately 450 new cases found each year, approximately 250 were male and 200 female. The male preponderance is seen from the diagram to be entirely due to the fact that nearly four times as many males as females were diagnosed over the age of 45. For females there was a peak of incidence at the age-group 15-24, with a sharp drop in incidence at higher ages. For males there was also a peak of incidence at the age-group 15-24, but the incidence rate, after falling to about 1 per 1 000 at age 35-44, remained steady for higher ages, in sharp contrast to the female pattern.

Classification of Disease at Diagnosis

Infectiousness. Within the sex- and age-groups the new notifications in 1952 and 1953 can be further divided according as they were sputum positive or negative when diagnosed. Fig. 2 shows the sex- and age-incidence of sputum positive cases with the numbers on which it is based given beneath.

Of the males forty-two per cent, and of the females thirty-one per cent had positive sputum; that is, were in an infectious condition when diagnosed. The reason for the sex difference is seen from the diagram to be due to the much higher incidence of positive sputum among older men than among older women. A comparison of Figs. 1 and 2 shows that although the peak of *total* incidence was in the female age-group 15-24, the peak of incidence of *infectious* cases was in the male group 55-64.

Stage of Disease. In 1952 and 1953 a total of 133 persons (83 males and 50 females) were not diagnosed until the disease was in a late stage. This is very unsatisfactory. Although the treatment now available can often prevent the death of these persons it will almost certainly be prolonged and costly, and they will always require medical supervision and suffer some disability. Moreover each advanced case has probably spread the disease to several other persons before being diagnosed.

Fig. 3 shows the age incidence of the three stages, early, intermediate and late, into which the disease was classified at diagnosis. For males the incidence of early cases showed a peak at ages 15-24 after which it

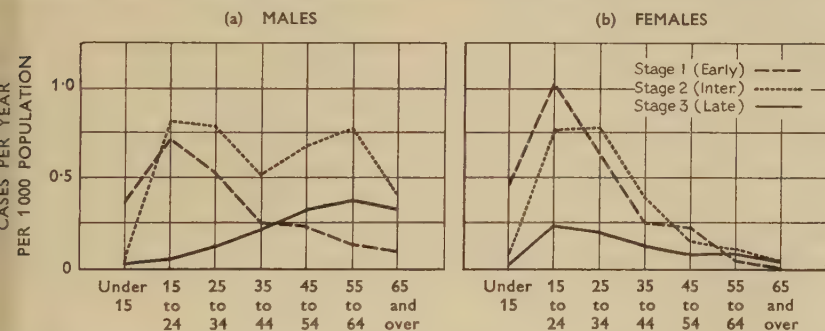


FIG. 3. Age-incidence of new cases of pulmonary tuberculosis in three stages of disease, Warwickshire 1952-53. The numbers of notifications on which the points in the diagrams are based are, in order:

Males:

Early:	42	41	37	21	15	6	4	total 166	} unclassified 21
Inter:	6	47	55	43	44	33	17	total 245	
Late:	3	3	8	18	21	16	14	total 83	

Females:

Early:	52	59	46	20	15	1	0	total 193	} unclassified 18
Inter:	9	44	55	31	10	6	2	total 157	
Late:	1	14	14	9	5	5	2	total 50	

fell steadily with age; the incidence of late cases showed a steady rise with age to reach its highest values at ages 45 and over. For females all three stages showed peaks at ages 15-24 and fell away at higher ages. The proportion of late to early cases was highest for males over 45. There was no difference between 1952 and 1953 in the proportion of cases diagnosed early or late. In future years this classification should enable us to judge the success of schemes directed at finding the early case. Unfortunately, the Warwickshire figures cannot be compared with others since no comparable sets of figures have been published.

Findings of Mass Radiography Surveys

During 1952 and 1953, 125 previously unknown cases of pulmonary tuberculosis were discovered by the Mass Radiography Unit during surveys in Warwickshire. This represents fourteen per cent of all new cases found in the county in these years. The Mass Radiography Unit

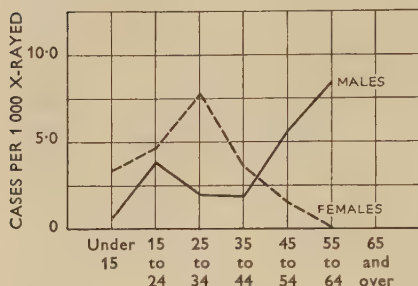


FIG. 4. Yield of Mass Radiography Surveys in Warwickshire, 1952-53. The numbers on which the points are based are, in order:

Numbers X-rayed:

Male: 4 725 3 075 4 029 3 184 2 847 711
total 18 571

Female: 4 366 4 901 3 086 2 499 2 054 351
total 17 257

Numbers of new cases:

Male: 3 12 8 6 16 6 total 51

Female: 15 23 24 9 3 0 total 74

of stage 2 and stage 3 cases shown in Fig. 3(a). This is reasonable since the cases which progress to stages 2 and 3 will probably go about unsuspected in the community for some considerable time before diagnosis.

Of the 125 cases found by mass radiography only four were in a late stage. This is striking confirmation that this method finds the early case.

Incidence of Pulmonary Tuberculosis in Different Districts of the County

The administrative County of Warwickshire contains within its boundaries extremely varied groups of population, ranging from widely scattered rural communities to densely populated industrial areas, with a number of good residential centres as well. Given these widely varied backgrounds one would expect variations in the incidence of pulmonary tuberculosis in the different districts.

To investigate the differences, the total number of new notifications for the five years 1949-53 has been found for each district, and used with the Registrar-General's mid-1951 population estimate, to calculate the average yearly rate of new notifications for each district. The total of five years' notifications was used for two reasons. One is because it gives larger, and therefore more reliable, numbers to work with. The other is because it tends to smooth out the effect of visits of the Mass

have kindly supplied us with the numbers X-rayed and the numbers of new cases found, divided into sex- and age-groups. The age-groups differ slightly at higher ages from the ones used so far. The rate of yield per 1000 X-rayed has been calculated and is shown in Fig. 4. This diagram is best studied in relation to Figs. 1 and 3. It will be seen that whereas the female yield-pattern is similar to the total-incidence pattern of Fig. 1, the male yield-pattern is unlike the total-incidence pattern. For males the yield-pattern, with its low peak at ages 15-24 and steady rise to its highest value at ages 55 and over, shows more similarity with the incidence-pattern

radiography Unit, which can put up the number of new notifications for a year in any district very considerably if a special effort is made here. Unfortunately, it was not possible to distinguish mass radiography cases accurately from other cases before 1952. Since most of the districts have been visited during the course of the five years this effect would be balanced out.

The results are shown in Table I. Even with five-year totals some of the districts have small numbers of notifications, and it was therefore thought desirable to make a statistical test of the difference between the rate for each district and the rate for the county as a whole.

TABLE I

Incidence of pulmonary tuberculosis in the sanitary districts comprising Warwickshire Administrative County

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
47 790	1	Sutton Coldfield MB	147	0.61	0.05	L	L
54 120	2	Nuneaton MB	369	1.36	0.07	H	H
25 080		Bedworth UD	199	1.58	0.11	H	H
23 580		Atherstone RD	98	0.83	0.08	—	—
45 850	3	Rugby MB	317	1.38	0.08	H	H
20 830		Rugby RD	80	0.77	0.09	—	—
37 950	4	Meriden RD	135	0.71	0.06	L	—
16 040		Tamworth RD	43	0.54	0.08	L	L
67 980	5	Solihull UD	231	0.68	0.04	L	—
36 520	6	Leamington Spa MB	153	0.84	0.07	—	—
15 550		Warwick MB	93	1.20	0.12	H	H
10 610		Kenilworth UD	40	0.75	0.12	—	—
22 700		Southam RD	47	0.74	0.11	—	—
19 940		Warwick RD	65	0.65	0.08	L	—
14 700	7	Stratford-on-Avon MB	49	0.67	0.10	L	—
13 260		Alcester RD	44	0.66	0.10	L	—
8 110		Shipston-on-Stour RD	22	0.54	0.12	L	—
20 390		Stratford-on-Avon RD	61	0.60	0.08	L	—
491 000		Total	2 193	0.89	—	—	—
365 950		Total excluding Nuneaton MB, Bedworth UD, Rugby MB	1 308	0.72	—	—	—

The columns are:

- (1) Population, Registrar-General's mid-1951 estimate.
- (2) Health administrative area.
- (3) Sanitary district.
- (4) Number of new cases of pulmonary tuberculosis, 1949-53.
- (5) Average yearly rate of new notifications per 1 000 of population.
- (6) Standard error of rate.
- (7) Rates in (5) significantly higher, H, or lower, L, than 0.89.
- (8) Rates in (5) significantly higher, H, or lower, L, than 0.72.

Consider, for example, the rate of 0.61 per 1 000 population for Sutton Coldfield MB, which is based on a total number of new notifications for the five years of 147 and a mid-1951 population of 47 790.

The standard error of this rate is

$$1\ 000 \sqrt{\left(\frac{\frac{147}{5 \times 47\ 790} \left(1 - \frac{147}{5 \times 47\ 790} \right) }{5 \times 47\ 790} \right)} \\ = \sqrt{(0.61/5 \times 47.790)} = \sqrt{0.0026} = 0.05$$

The difference between the rate of Sutton Coldfield and the county rate is $0.89 - 0.61 = 0.28$, which is more than twice the standard error of 0.05 and is therefore statistically significant. (The county rate is assumed to be without error.)

The standard errors have been calculated in this way for all district rates and are shown in the table. The districts with rates significantly higher or lower than the county average of 0.89 are marked. There are four districts with significantly high rates, and no fewer than nine with significantly low rates. On examination it appears that three districts have exceptionally high rates, namely Nuneaton MB, Bedworth UI and Rugby MB. If an average is calculated by excluding these districts and the comparisons made with this new average, one district remains significantly high, namely Warwick MB, but only two districts remain significantly low, namely Sutton Coldfield MB and Tamworth RD.

The main conclusion, then, is that there are three or four 'black spots' for pulmonary tuberculosis in the county. They are Nuneaton and Bedworth, which are adjacent industrial and mining towns, very densely populated with a good deal of poor housing; Rugby with expanding industries superimposed on an old market town; and, to a lesser degree, Warwick, which is the county town but does contain some industry and some poor housing conditions. The rest of the county is fairly uniform with the exception of borderline significantly low rates in Sutton Coldfield, a good residential district bordering on Birmingham and Tamworth RD.

The sex- and age-incidence and the numbers found in each stage of the disease were very similar for the towns with high rates of notification as for the towns with lower rates.

Application of the Statistical Findings

Much more effective treatment has recently become available for pulmonary tuberculosis than was available in the past. This is reflected in the death rates, which have fallen dramatically in recent years, from a level in Warwickshire of about 0.40-0.45 per 1 000 population between 1935 and 1948 to 0.19 in 1952 and 1953. These new methods of treatment mean, firstly, that it is of the greatest benefit to the individual sufferer from the disease that he should be diagnosed, and in as early a stage as possible. Secondly, that the control and even elimination of the disease are now feasible, providing that each affected person is diagnosed as soon as possible, before he has infected others.

If limitless resources were available and compulsion were possible, early diagnosis could be achieved by X-raying the entire population at frequent intervals. Since only limited facilities exist, and persuasion and publicity have to be used instead of compulsion, it is important that efforts should be directed to the sections of the population where the greatest reservoir of undiagnosed infectious cases lies. This analysis has shown that for Warwickshire certain towns need more attention than others, and that the group of males over 45 needs more attention than the other age- and sex-groups.

It is realised that there are other factors which should be taken into account, particularly occupation. Unfortunately, we have not as yet sufficient information to be able to make other analyses.

My thanks are due to Dr S. W. Savage, County Medical Officer of Health for Warwickshire, for his help and advice.

MULTIPLE REGRESSION BETWEEN VARIABLES MEASURED AT THE POINTS OF A PLANE SHEET

GEOFFREY H. JOWETT

University of Sheffield

A correspondent submitted to *Applied Statistics* a question concerning the analysis of the results of measurements made over a plane sheet of steel. The issues raised were too big to be dealt with in the short space of a Question and Answer and Dr Jowett has written this article. The article is concerned with the removal of trends in determining regressions and gives a test for patchiness in residuals. It also deals with analysis after trend or patchiness has been removed.

Question

The question postulated was the following.

'We are interested in the factors affecting the variation of a property of sheet steel. In order to examine how it varied over the area of a sheet we cut the sheet into 130 strips, by nine transverse cuts and twelve longitudinal cuts. We then measured the property on each of the small strips. At the same time we measured four other physical, chemical, and metallurgical characteristics of the material. We have carried out a multiple regression analysis of the property with position and with the other characteristics, and would appreciate your comments on the statistical interpretation of the results.

'The full analysis was carried out with twenty-six variables, X_1 to X_{26} , twenty-one of which were orthogonal position terms.

X_1 = linear orthogonal term H_1 for the horizontal variation fitted to the means for the vertical strips,

X_2 = quadratic orthogonal term H_2 for the horizontal variation up to $X_6 = H_6$,

X_7 = linear orthogonal term V_1 for the vertical variation, fitted to the means for the horizontal strips,

up to $X_{12} = V_6$,

$X_{13} = H_1 V_1$,

$X_{14} = H_1 V_2$,

$X_{15} = H_1 V_3$,

up to $X_{20} = H_3 V_2$,

$X_{21} = H_3 V_3$,

X_{22} = first characteristic C_1 ,

X_{23} = second characteristic C_2 ,

X_{24} = third characteristic C_3 ,

X_{25} = fourth characteristic C_4 , and

$X_{26} = C_4^2$.

'Now the result of this analysis (analysis 1) was that the property was related to six of the position terms and to only one characteristic C_1 (at the 1 in 100 level). The characteristics and position variables together accounted for 76 per cent of the total variation. Our interpretation of this was that there was some other characteristic (or characteristics) which varied over the area of the sheet and which affected the measured property. The property was considered to be independent of the characteristics C_2 , C_3 , C_4 , and C_4^2 .

'Because of the labour of inverting such a large matrix a preliminary analysis was made considering only the three position terms with a highly significant (at the 1 in 1 000 level) correlation with the measured property, and the characteristics C_1 , C_2 , C_3 , C_4 , and C_4^2 . This analysis (analysis 2) indicated that the property depended on position, and also on C_4 and C_4^2 (both of these latter at the 1 in 20 level). In this case the measured property appeared to be independent of C_1 , C_2 , and C_3 . Our interpretation of this analysis, in the light of the more detailed one, was that one or more of the position terms which should have been included varied in a similar manner to C_4 and C_4^2 , and in a contrary manner to C_1 .

'If the effect of position could be ignored (analysis 3), then the final property depended on C_1 , C_4 , and C_4^2 (at the 1 in 1 000 level), and on C_3 (at the 1 in 20 level). The picture given by this relationship is easier to interpret, but is it the true picture? These characteristics together accounted for only 43 per cent of the total variation.

'The outer edge strips were discarded, and this relationship remained the same (statistically) for the central area as for the whole area (analysis 4).

'The queries I wish to pose are:

'Do you agree with the interpretation of analysis 1, that there is some one or more characteristic of prime importance varying with position and consequently affecting the final property?

'How is it that, if the variation from position to position is ignored, the characteristics C_1 , C_3 , C_4 , and C_4^2 are related to the final property (analyses 3 and 4)? The characteristics C_1 , C_2 , C_3 , and C_4 vary systematically over the area of the sheet (each has its own pattern with considerable random fluctuations about it). Pattern or no pattern, if these characteristics were related to the measured property, should not they show up in analysis 1 as being the prime movers rather than the variation from position to position? That is to say, should not the final property tend to follow these fluctuations about the pattern, and prefer to be related to them rather than to the position terms? Is it possible that the property is a function of C_1 , C_2 , C_3 , and C_4 which is not readily approximated by the linear terms of C_1 , C_2 , C_3 , and C_4 and the quadratic term C_4^2 ?

Introduction to Answer

It is easiest to start with a discussion of the interpretation of an

analysis, like analysis 3, of the question and to leave aside the complications introduced by the inclusion of position variables till later. Moreover, since problems of this kind are of wide occurrence, I shall use general terms to begin with, and leave the answering of the question itself till the end. Many of the considerations that affect the analysis of data of this kind also apply to time series or other linear series such as properties measured along the length of steel strip or rod.

Suppose we have records of, say, two 'cause' variables x_1 and x_2 and an 'effect' variable y , of which the values are given at an evenly spaced network of points over a sheet, as described in the question; and suppose that we carry out a regression analysis in the ordinary way, arriving at the fitted equation

$$Y = a + b_1x_1 + b_2x_2 \quad \dots (1)$$

If the residual variation in y or, in other words, that part of its variation which is not due to the cause variables, is random over the sheet, it is permissible to compute the standard errors of b_1 and b_2 in the usual way and to use them to test the significance or measure the sampling accuracy of b_1 and b_2 . Two problems arise here: the first is the checking of randomness in the residuals, the second is the interpretation of b_1 and b_2 in terms of cause and effect.

To check randomness in the residuals it is necessary to estimate them, using the regression equation; for values x_{1i} , x_{2i} , y_i at a point i of the sheet, an estimate of the residual e_i is given by the equation

$$e_i = y_i - \hat{Y}_i = y_i - a - b_1x_{1i} - b_2x_{2i} \quad \dots (2)$$

These estimated residuals will not be completely random, even in the most favourable circumstances, because of the fitting of the constants of regression; for example, their sum is necessarily exactly zero. This type of departure from randomness, however, is of very minor importance compared with the other possible types, and may be neglected; it is very unobtrusive, anyhow. More serious are departures which fall into one or other of these two types: (i) two-dimensional trends over the sheet, e.g. a tendency to fall near the edges and to be high in the middle, or to tend to increase from one side to the opposite side, (ii) correlation of neighbouring residuals, which often shows itself as a kind of patchiness over the sheet; in one small region residuals may tend to be mostly high, with a transition to lowness in another region. The two states of affairs are illustrated in Figs. 1(a) and 2(a), which show contour maps of a typical trend and the non-random part of a typical piece of patchiness, the contours being for constant size of residual, and the rectangles representing the sheet.

Investigation of Trend or Patchiness in Residuals

Trend or patchiness in residuals could be investigated by mapping the residuals, like heights in surveying, and interpolating contours as cartographers do; but often the residuals have so much random

variation on top of this trend or patchiness that this method is too difficult. It is easier to plot the residuals in sections, as in Figs. 1(b) and 2(b), and assess the state of affairs from these.

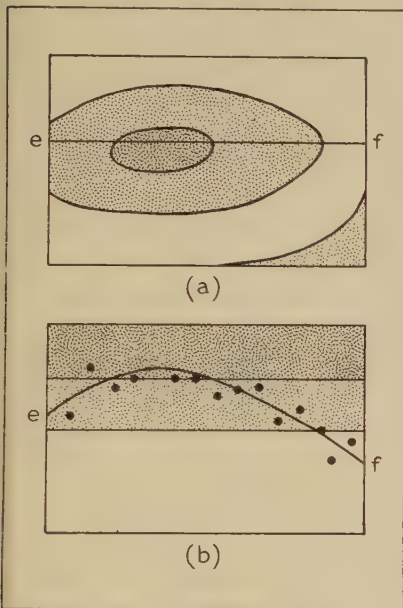


FIG. 1. (a) Contour map showing two-dimensional smooth trend. (b) Residuals along the section *ef* of the surface in (a).

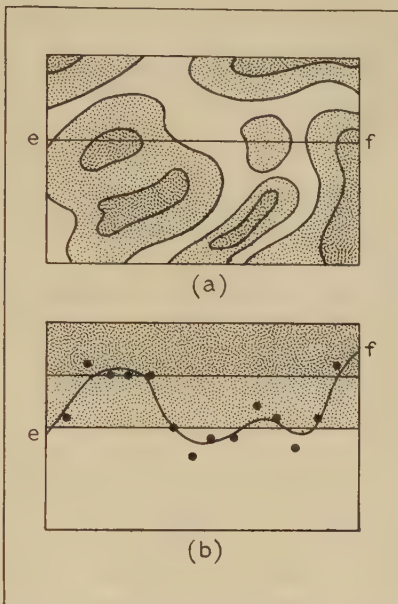


FIG. 2. (a) Contour map showing two-dimensional smooth patchiness. (b) Residuals along the section *ef* of the surface in (a).

A preliminary graphical exploration of this kind may be sufficient in itself, or the investigator may require, in doubtful cases, a more formal test of trend or patchiness. If he suspects that a trend exists, he might fit enough polynomial terms to reproduce what he suspects is the general topography (e.g. for Fig. 1(a) terms u , u^2 , v , v^2 would probably do, u and v being co-ordinates of position within the sheet) and test whether they account together for a significant proportion of the variation; for this purpose it is best to be as economical as possible in the use of polynomial terms, or the power of the test tends to be reduced. On the other hand, if he suspects patchiness, he might test for its existence with a statistic such as

$$Av(e_i - E_i)^2$$

where E_i is the mean of the eight residuals which are adjacent to the residual e_i , forming a square round it, and where the averaging operation is taken over all the cases for which the quantity to be averaged can be calculated (i.e. for all e_i except those at the edge of the sheet). If patchiness exists, this statistic will tend to be smaller than the mean square of the e_i themselves, since in patchiness values tend to resemble those round about; if variation is random, it will tend to be slightly larger, because of the added variation of E_i and, as Appendix I shows,

will actually average nine-eighths of that mean square. Hence the precise test of patchiness consists of calculating the ratio

$$U = (8/9) \text{Av}(e_i - \bar{E}_i)^2 / \text{Av}(e_i^2) \quad \dots (3)$$

and testing the significance of its difference from an expected value of unity by means of its standard error, which for large samples and normally distributed e_i is shown in Appendix I to be given approximately by the formula

$$\sigma_U^2 = 179/324n \quad \dots (4)$$

where n is the number of terms averaged. For one-dimensional series we could test for serial correlation in the residuals (which is the equivalent of this patchiness) by the method given by Mr Durbin and Mr Watson in a paper in *Biometrika*.¹ Trend and patchiness are not well-distinguished ideas, for an apparent trend in a small section of plate may show up as a patch in the whole. For test purposes patchiness may be taken to begin where the pattern of variation becomes too complicated to be represented by any polynomial of reasonably low degree.

Interpretation of Regression Coefficients when Residuals are Random

Although standard errors of the coefficients b_1 and b_2 may be calculated in the usual way if trend and patchiness appear to be absent, there remains a difficulty in interpreting the coefficients in terms of cause and effect in addition to the difficulties normally experienced for random data. It may happen, for instance, that a third important cause variable x_3 , omitted from the analysis, is fortuitously correlated with x_1 merely because its pattern of variation over the sheet happens to be somewhat similar to that of x_1 . This is called a *nonsense correlation* and can easily happen when there are trends in the variables, because there are only a few shapes which such trends can take—in contrast to the wide variety of random patterns of variation—and chance coincidence becomes quite likely. If this happens, x_1 will have its apparent effect inflated by the concomitant effect of x_3 and, since multiple regression works on the principle of ‘to him that hath shall be given’, will take the lion’s share of the credit for the joint effect of x_1 and x_2 if these happen to be correlated, and x_2 has no similar hidden allies to back it up in the fight for recognition of its importance. Even when there is no important single variable omitted, this kind of difficulty can be caused by a complex of minor variables working in step with each other and with x_1 , and it is this which makes the interpretation of multiple regression between time and spatial series such a hazardous business.

Regression Analysis when Residuals are not Random

When trend exists, it may be removed from all three variables and the regression analysis may be carried out on what is left. This may be done by including as further independent variables in the regression enough polynomial terms to be sure of fitting the trend adequately in

all three cases. This is what the questioner has tried to do. The regression coefficients b_1 and b_2 are tested for significance in the usual way, which may require much calculation because of all the new independent variables; and then the polynomial coefficients are henceforth disregarded, interpretation being confined to b_1 and b_2 and making the usual form.

Alternatively or when patchiness exists (the same method will serve for both), we may remove it by subtracting some suitable moving average, which should be the same for each, from the values of all three variables x_1 , x_2 , and y . I have discussed this technique for time series in another paper,² and it is impossible to summarise all its important aspects in a few words. It is possible, however, to describe a simplified version of the technique which will serve well enough in many practical applications.

From each value of any variable we subtract the average of that value and the eight adjacent values forming a square around it. At the edges of the sheet there may not be eight adjacent values, but since this is to be a large-sample test, it suffices to use such values as there are. The regression between the residuals is then carried out in the usual way, except that there is no need to subtract corrections for the mean from sums of squares and products. In calculating standard errors, however, it is necessary to multiply the estimates as usually calculated by a factor of 1.13, to take account of the disturbance caused by the moving average in the randomness of the residuals from regression. This factor is justified in Appendix II. The method itself is valid provided that the variables can be regarded as made up of either (i) smooth trend + random variation or (ii) smooth patchiness + random variation, and that the smooth trend or smooth patchiness is roughly linear over any area covered by a square of nine points. The method will be reasonably efficient if there is considerable random variation left after the removal of the trend or smooth patchiness.

In these trend-removal methods of regression there is always a risk that some important information may be thrown out with the trends, particularly if these are too closely fitted. Yet it is often better to use a less efficient method which can be trusted rather than a more sensitive method which may break down without giving any evidence of it.

Comment on the Question as Posed

It seems likely that the significance tests of the questioner's analysis 3 are untrustworthy because of the systematic patterns of variation in all the variables concerned. If the patterns are so marked as to be capable of giving apparent significance to the position variables in analysis 1, apparent significances (judged by the standard test) such as those in analysis 3 would be quite likely to arise merely from chance incidences in the forms of the patterns.

If the picture of relationships presented by analysis 3 happens by good fortune to be correct, the trend removal implicit in analysis 1

will, if adequate, leave deviations between which these relationships may certainly be expected to tend to persist, and the test involving regression coefficients on $C_1 \dots C_4$, C_4^2 will be statistically valid. Unfortunately this does not imply that these coefficients need remain significant; the trend removal may have left behind too little information for their significance to be established even in spite of their considerable residual variation about their patterns. Hence the significant results in analyses 1 and 3 are not mutually inconsistent, nor necessarily inconsistent with the characteristics $C_1 \dots C_4$ being prime movers. The significance of the regression coefficients for the position variable in analysis 1 means simply that correlated trends are being removed and the correlation may be fundamental or fortuitous, for the standard tests do not discriminate. Hence it is not necessary on these grounds to postulate the existence of another characteristic of prime importance nor to conclude that the property is a more complicated function of $C_1 \dots C_4$ than has been assumed. On the other hand, either of these possibilities could be sufficient to account for the facts, and it might be worth while to scrutinise the possibility of parabolic regression in C_1 , C_2 , and C_3 by scatter diagrams.

The fact that analysis 1 explains so much more of the variation of the effect variable than analysis 3 suggests that a considerable part of its variation is neither strictly random nor dependent on $C_1 \dots C_4$, and has sufficient broad trend-like features to be capable of making a substantial contribution to the 'effects' of the position variables and thus giving them more pulling power in the general share-out of credit. It would be rash, however, to argue that some cause variable with status comparable to that of $C_1 \dots C_4$ must necessarily have been overlooked, because non-random variation in spatial and linear series is often just as much a consequence of 'random causes' as strictly random variation. For example, work-hardening through randomly distributed stresses would tend to be distributed in a patchy manner, and the smoothing effect of, say, partial annealing might conceivably be enough to turn this patchiness into a 'position effect'. The questioner's own interpretation, based as it is on evidence which is rather negative in character, is thus hardly a safe one.

There is clearly room for divergence of opinion in the interpretation of the results given. Some of this might be removed by re-scrutiny of the data in the spirit of this article. The laboriousness of the polynomial method of approach in this case makes one wonder if a moving average method might have been easier and at least as informative.

It is doubtful if any useful interpretation can be given to analysis 2. The rejection of polynomial terms can only really be justified on the grounds of ineffectiveness in removing trend; with such a high significance level, some which were effective either individually or in combination might have been rejected unjustifiably.

I am indebted to Dr O. L. Davies for his opinion on the problem raised in the question.

APPENDIX 1: TEST FOR PATCHINESS

The statistic U given by equation (3) may also be defined as follows:

$$U = (8/9) \Sigma (e_i - E_i)^2 / \Sigma e_i^2 \quad \dots (6)$$

Large-sample expectation and variance are given by the formulae

$$E(U) \sim E[\frac{8}{9} \Sigma (e_i - E_i)^2] / E(\Sigma e_i^2) \quad \dots (7)$$

$$\text{var}(U) \sim \text{var}[\frac{8}{9} \Sigma (e_i - E_i)^2] - 2E(U) \text{cov}[\Sigma e_i^2, \frac{8}{9} \Sigma (e_i - E_i)^2] + [E(U)]^2 \text{var}(\Sigma e_i^2) / [E(\Sigma e_i^2)]^2 \dots (8)$$

Since e_i and E_i are independent, we have the following relations, of which (10) and (11) hold only if e_i is normally distributed:

$$E(e_i - E_i)^2 = \sigma^2(1 + \frac{1}{8}) \quad \dots (9)$$

$$\text{cov}[e_i^2, (e_j - E_j)^2] = \begin{cases} 2\sigma^4 & \text{if } i = j \\ \sigma^4/32 & \text{if } i \text{ is laterally or diagonally next to } j \\ 0 & \text{otherwise} \end{cases} \quad \dots (10)$$

$$\text{cov}[(e_i - E_i)^2, (e_j - E_j)^2] = 2[\text{cov}\{(e_i - E_i), (e_j - E_j)\}]^2 \quad \dots (11)$$

Using (9) in (7), we see that

$$E(U) \sim 1 \quad \dots (12)$$

In evaluating the first two terms on the right-hand side of (8) we note that the partial summation over j gives the same answer for every i . Hence

$$\text{cov}[\Sigma e_i^2, \frac{8}{9} \Sigma (e_i - E_i)^2] \sim \frac{8}{9} (2 + \frac{8}{32}) n \sigma^4 = 2n\sigma^4 \quad \dots (13)$$

For given i , the covariances on the right-hand side of (11) have non-zero values only when i and j are laterally or diagonally separated by two intervals at most. They follow the scheme:

$$\frac{1}{64} \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & -14 & -12 & -14 & 2 \\ 3 & -12 & 72 & -12 & 3 \\ 2 & -14 & -12 & -14 & 2 \\ 1 & 2 & 3 & 2 & 1 \end{bmatrix} \sigma^4 \quad \dots (14)$$

Hence the sum of squares of these covariances is $827n\sigma^4/256$. Hence

$$\text{var}[\frac{8}{9} \Sigma (e_i - E_i)^2] \sim 827n\sigma^2/324 \quad \dots (15)$$

$$\text{var}(\Sigma e_i^2) = 2n\sigma^4 \quad \dots (16)$$

Substituting in (8), we have the final result:

$$\text{var}(U) \sim 179/324n$$

APPENDIX II: CORRECTING FACTOR FOR STANDARD ERRORS IN REGRESSION BETWEEN SERIES FROM WHICH MOVING AVERAGE TRENDS HAVE BEEN REMOVED

When a variable which is random at a network of points over a sheet is subjected to a trend-removing operation, using the moving average described in the text, the resulting values are no longer strictly random, but have spatial correlations as shown below:

.	.	.	.	.	.	.	.	.
.	.	1/72	1/36	1/24	1/36	1/72	.	.
.	.	1/36	-7/36	-1/6	-7/36	1/36	.	.
.	.	1/24	-1/6	1	-1/6	1/24	.	.
.	.	1/36	-7/36	-1/6	-7/36	1/36	.	.
.	.	1/72	1/36	1/24	1/36	1/72	.	.
.	.	.	.	.	.	.	.	.

.....(17)

Thus, for example, laterally adjacent values have a correlation of $-1/6$, diagonally adjacent values a correlation of $-7/36$, and so on.

If two such variables with correlation ρ are treated in this way, the spatial lag correlations between them become proportional to those shown above, the constant of proportionality being ρ .

If x_1 and x_2 are two cause variables and y is an effect variable (all stationary), the large-sample error variances and covariances of the regression coefficients b_1 , b_2 are given by the matrix equation

$$\begin{pmatrix} \text{var } b_1 & \text{cov } b_1 b_2 \\ \text{cov } b_1 b_2 & \text{var } b_2 \end{pmatrix} \sim \frac{1}{n} \begin{pmatrix} \text{var } x_1 & \text{cov } x_1 x_2 \\ \text{cov } x_1 x_2 & \text{var } x_2 \end{pmatrix}^{-1}$$

$$\begin{pmatrix} A \text{ var } x_1 & \text{var } \epsilon & B \text{ cov } x_1 x_2 & \text{var } \epsilon \\ B \text{ cov } x_1 x_2 & \text{var } \epsilon & C \text{ var } x_2 & \text{var } \epsilon \end{pmatrix} \begin{pmatrix} \text{var } x_1 & \text{cov } x_1 x_2 \\ \text{cov } x_1 x_2 & \text{var } x_2 \end{pmatrix}^{-1}$$

..... (18)

where ϵ is the residual from the true regression, A is the sum of products of corresponding spatial correlations for x_1 and ϵ , B is the corresponding quantity for x_2 and ϵ , and C is the ratio of the sum of products of lag correlations of x_1 and x_2 with the spatial correlations of ϵ to the correlation of x_1 and x_2 .

For the special case described above, A , B , and C are all equal to the sum of squares of the figures in the table, which is $6616/5184$; this factor may therefore be taken out of the matrix product, which then takes the form that would be appropriate if x_1 , x_2 , and ϵ varied randomly. Thus the standard errors of b_1 and b_2 as ordinarily calculated should be multiplied by the square root of this factor, which is 1.13 . The application of this result in the manner described in the text presupposes that the nine-term moving average removes all the non-random variation from x_1 , x_2 , and ϵ .

For one-dimensional series, the use of the corresponding three-term moving average is valid but rather less satisfactory, since it is not such a powerful smoothing operator and tends to throw away rather a lot of information; none the less, it can be used, and requires a correcting factor for standard errors of 1.39.

REFERENCES

- DURBIN, J., and WATSON, G. S. (1950). 'Testing for serial correlation in least squares regression', *Biometrika*, **37**, 409.
- LOWETT, G. H. (1955). 'Least squares regression analysis for trend-reduced time series', *J.R. Statist. Soc.*, B (in the press).

CALCULATION BY ACCUMULATION

P. A. DANIELL and F. D. K. LIDDELL

Mr Daniell and Mr Liddell describe a previously known procedure for calculating various quantities from frequency distributions, and show how time can be saved by using Hollerith equipment.

MANY of the calculations normally carried out on a frequency distribution can be done by repeated summation, provided that the class interval of the distribution is constant. This paper discusses the method which, of course, is not new.¹ It has already proved its use in many places, but we think that there may be people to whom it would be useful and who are still unaware of it. By its nature it can be handled either without machines or with any simple adding machines; its most efficient application is when the Hollerith sorter and tabulator are available.

To illustrate the method we use the distribution, given in Table I, of seventy-seven wage-earners classified according to their weekly

TABLE I

Weekly earnings of seventy-seven wage-earners

Earnings, shillings	Number of Men
98	5
100	17
102	27
104	25
106	0
108	3
Total ..	77

earnings. For the first use of the method we form the cumulative total of the numbers of men and then accumulate these totals again. Table II shows the results of this process. Where there is a zero frequency, as here at earnings of 106 shillings, the cumulative total has still to be entered in the appropriate place and it must be added in the second accumulation.

TABLE II

Accumulations of distribution of wage-earners

Earnings, shillings (x)	Numbers of Men		
	Actual	Accumulated	Accumulated again
98	5	5	5
100	17	22	27
102	27	49	76
104	25	74	150
106	0	74	224
108	3	77	301
Total ..	77	301	783

Now let h be the class interval (in this case 2 shillings),

k be the number of values of the variate (6),

x_1 be the first value of the variate (98),

N be the total frequency (77, checked by the last term in the first accumulation),

A be the total of the first accumulation (301, checked by the last term in the second accumulation), and

B be the total of the second accumulation (783).

Then we can find the total earnings of these men by use of the simple formula

$$\Sigma x = (x_1 + hk)N - hA$$

which in our case becomes

$$\begin{aligned}\Sigma x &= (98 + 12) \times 77 - 2 \times 301 \\ &= 7\,868 \text{ shillings}\end{aligned}$$

If we wish to find the variance, the sum of squares about the mean is found from

$$\Sigma(x - \bar{x})^2 = h^2(2B - A - A^2/N)$$

and in our case this is

$$\begin{aligned}\Sigma(x - \bar{x})^2 &= 4[2 \times 783 - 301 - (301)^2/77] \\ &= 353.45\end{aligned}$$

Assume that we wish to estimate the cost of raising the earnings of over-paid workers to various figures. This method of calculation affords a rapid solution. To get our estimate we add to the present figures the difference between our second and first accumulations at the new level of minimum earnings, multiplied by the class interval.

Thus if in our example we were to consider raising the earnings of the twenty-two men who earn less than 102 shillings to that figure, the increase in our wages bill is

$$2 \times (76 - 49) = 54 \text{ shillings}$$

and the new total of earnings would be

$$7\,868 + 54 = 7\,922 \text{ shillings}$$

The second use of the method (slightly adapted from that given in references 2 and 3) is not quite so simple for pencil-and-paper calculation but is very easy with Hollerith machinery. It is suitable for finding sums of squares and of cross-products. As an example, we shall consider the data of Table I, further subdivided according to the tax payments of the men.

TABLE III

Numbers of wage-earners with different weekly earnings and making different tax payments

Earnings, shillings (x)	Tax Payments, shillings (y)					Total
	10	15	20	25	30	
98	0	3	1	1	0	5
100	0	5	9	2	1	17
102	1	7	10	7	2	27
104	0	9	11	5	0	25
106	0	0	0	0	0	0
108	0	0	1	0	2	3
Total ..	1	24	32	15	5	77

TABLE IV

Accumulation of distribution of wage-earners

Earnings, shillings	Number of Men	Total Earnings, shillings		Total Tax Payments, shillings	
		Of these Men	Accumulated	Of these Men	Accumulated
98	5	490	490	90	90
100	17	1 700	2 190	335	425
102	27	2 754	4 944	550	975
104	25	2 600	7 544	480	1 455
106	0	0	7 544	0	1 455
108	3	324	7 868	80	1 535
Total ..	77	7 868	30 580	1 535	5 935

Now for each value of earnings we list the number of men, their total earnings, and their total tax payments. Thus the total earnings of the men earning 98 shillings each are 490 shillings and they pay in tax a total of $3 \times 15 + 1 \times 20 + 1 \times 25 = 90$ shillings. Then we find the cumulative totals of the last two items and sum these columns. Table IV gives the results.

Now let C be the total earnings (7 868, checked by the last term of the accumulation),

D be the total of the accumulations of earnings (30 580).

E be the total tax payments (1 535, checked by the last term of the accumulation), and

F be the total of the accumulations of tax payments (5 935).

Then the total earnings of the men are found direct ($C = 7\ 868$) and the sum of squares is found from

$$\Sigma x^2 = C(x_1 + hk) - hD$$

which in this case

$$\begin{aligned} &= 7\ 868 \times (98 + 12) - 2 \times 30\ 580 \\ &= 804\ 320 \text{ shillings} \end{aligned}$$

and the sum of squares about the mean is

$$\Sigma(x - \bar{x})^2 = \Sigma x^2 - (\Sigma x)^2/N$$

$$\begin{aligned} \text{which becomes } \Sigma(x - \bar{x})^2 &= 804\ 320 - (7\ 868)^2/77 \\ &= 353\cdot45 \text{ as before} \end{aligned}$$

The total tax payments are found direct ($E = 1\ 535$) and the sum of cross-products is

$$\Sigma xy = E(x_1 + hk) - hF$$

$$\begin{aligned} \text{and this } &= 1\ 535 \times (98 + 12) - 2 \times 5\ 935 \\ &= 156\ 980 \text{ shillings} \end{aligned}$$

and the sum of cross-products about their means is

$$\Sigma(x - \bar{x})(y - \bar{y}) = \Sigma xy - (\Sigma x)(\Sigma y)/N$$

which, in the particular case,

$$= 156\ 980 - (7\ 868) \times (1\ 535)/77 = 130\cdot91$$

If a Hollerith punched card exists for each man, all this totalling can be done, after one sorting, on a single run through the tabulator. Indeed, as many 'dependent' variables can be treated at once as the particular tabulator can manage. An 11-wheel counter can be split to give two or even three variables, so that a tabulator with six counters could be used to find in a single run the sums of cross-products of one variable with itself and at least three others.

Even if the range of the 'independent' variable is very large, the process can be handled in a few minutes. If it is desired to reduce the time spent on the tabulator to a minimum, the data can be grouped. The only difficulty with a fine interval of the 'independent' variable is the possibly large number of values with zero frequencies. A card must be inserted into the sorted pack for each such blank cell. The most convenient method of dealing with this appears to be that given by Hartley³ of using two packs of 'space' cards—cards of one colour being punched x in odd columns and y in even columns, and cards of a second colour being punched y in odd columns and x in even columns. Whenever one of these cards is sensed in the tabulator, control will be broken and another total will be entered, but it will have no other effect. Any subsequent sort of the cards will automatically throw the 'space' cards into the x and y boxes of the sorter.

The time saved by this method compared with calculation by desk machine depends greatly on the size of the problem. If cards are already punched, there will be slight savings even when dealing with small packs of cards and only two or three variables. For larger jobs the method is efficient enough to justify punching cards *de novo* for the particular exercise. In a recent exercise, dealing with four variables on packs of only about twenty cards, already punched, the saving in time to produce the raw sums of squares and cross-products was estimated at about a third.

APPENDIX: DERIVATION OF FORMULAE

The direct derivation of the general formulae used in the body of the text is not difficult, but involves tedious algebra. They can be derived indirectly by transferring to an origin at the first value of the variate and to a scale with class interval unity. For the first use of the method the data become as follows:

Variate		Frequency		
Original	Trans- formed	Of Variate	Accumulated	Accumulated again
x_1	0	f_0	f_0	f_0
$x_1 + h$	1	f_1	$f_0 + f_1$	$2f_0 + f_1$
...	...	...	...	...
$x_1 + rh$	r	f_r	$f_0 + f_1 + \dots + f_r$	$(r+1)f_0 + rf_1 + \dots + f_r$
...	...	...	...	...
$x_1 + (k-1)h$	$k-1$	f_{k-1}	$f_0 + f_1 + \dots + f_r + \dots + f_{k-1}$	$kf_0 + (k-1)f_1 + \dots + (k-r)f_r + \dots + f_{k-1}$
Total ..		N	A	B

In the transformed variate, ξ , say,

$$\Sigma(\xi) = \sum_{r=0}^{k-1} (r \cdot f_r) \quad \dots (1)$$

$$\Sigma(\xi^2) = \sum_{r=0}^{k-1} (r^2 \cdot f_r) \quad \dots (2)$$

The totals A and B are seen to be

$$\begin{aligned} A &= kf_0 + (k-1)f_1 + \dots + (k-r)f_r + \dots + f_{k-1} \\ &= \sum_{r=0}^{k-1} \{(k-r)f_r\} \\ &= \sum_{r=0}^{k-1} (k \cdot f_r) - \sum_{r=0}^{k-1} (r \cdot f_r) \\ &= kN - \sum_{r=0}^{k-1} (r \cdot f_r) \quad \dots (3) \end{aligned}$$

d

$$\begin{aligned} &= f_0[1+2+\dots+r+\dots+k] + f_1[1+2+\dots+r+\dots+(k-1)] \\ &\quad + \dots + f_r[1+2+\dots+(k-r)] + \dots + f_{k-1} \\ &= \sum_{r=0}^{k-1} \{f_r[1+2+\dots+(k-r)]\} \\ &= \sum_{r=0}^{k-1} \left\{ f_r \frac{(k-r)(k-r+1)}{2} \right\} \\ &= \sum_{r=0}^{k-1} \left\{ \frac{1}{2}k(k+1)f_r - \frac{1}{2}(2k+1)r \cdot f_r + \frac{1}{2}r^2 \cdot f_r \right\} \\ &= \frac{1}{2}k(k+1)N - \frac{1}{2}(2k+1) \sum_{r=0}^{k-1} (r \cdot f_r) + \frac{1}{2} \sum_{r=0}^{k-1} (r^2 \cdot f_r) \quad \dots (4) \end{aligned}$$

Then, from (1), (2), (3), and (4):

$$\Sigma(\xi) = kN - A \quad \dots (5)$$

d

$$\begin{aligned} \Sigma(\xi^2) &= 2B - k(k+1)N + (2k+1)(kN - A) \\ &= 2B - (2k+1)A + k^2N \end{aligned}$$

$$\text{hence } \Sigma(\xi - \bar{\xi})^2 = 2B - (2k+1)A + k^2N - (kN - A)^2/N$$

$$= 2B - A - A^2/N \quad \dots (6)$$

Now transforming back from ξ to x , we note

$$\bar{x} = x_1 + h\bar{\xi}$$

and so

$$\Sigma(x) = N x_1 + h \cdot \Sigma(\xi)$$

$$= N(x_1 + hk) - hA \quad [\text{from (5)}]$$

and

$$\Sigma(x - \bar{x})^2 = h^2 \Sigma(\xi - \bar{\xi})^2$$

$$= h^2(2B - A - A^2/N) \quad [\text{from (6)}]$$

The other formulae follow from similar arguments.

REFERENCES

- ¹ AITKEN, A. C. (1947). *Statistical Mathematics*. Fifth edition, p. 40. Oliver and Boyd, London and Edinburgh.
- ² COMRIE, L. J., *et al* (1937). 'Application of Hollerith equipment to an agricultural investigation', *J.R. Statist. Soc.*, Supp. **4**, 210.
- ³ HARTLEY, H. O. (1946). 'The application of some commercial calculating machines to certain statistical calculations', *ibid.*, Supp. **8**, 154.

A SIMPLE NON-PARAMETRIC GRAPHICAL TEST FOR THE SIGNIFICANCE OF A DIFFERENCE

WLADYSŁAW Z. BILLEWICZ

*External Staff, Medical Research Council,
Midwifery Department, University of Aberdeen*

Mr Billewicz illustrates by some data on birth weights a simple method that should have application in many practical situations. The method avoids a complicated regression analysis and presents the results in a way that can be appreciated by people who are not statisticians.

Introduction

Most of the questions arising in the statistical analysis of an extensive collection of records can be approached by several different methods. Each statistician has his personal preferences but, in general, the method of choice is that which answers the question or questions most clearly, which makes efficient use of all the available material, and which is economical in terms of time and effort. The technique which is most efficient in a purely statistical sense may involve very great computational labour and may therefore in practice be better replaced by a method that is much quicker and simpler at the price of a small loss of information. The choice will be further guided by the importance of not only answering the leading question but also revealing as much as possible about various side issues. Finally, a point frequently overlooked, the statistician should select a method that is understandable, at least in general outline, to his non-statistical colleagues and the readers of his report; only so can he ensure a critical appreciation of the results and limitations of his analysis.

In medical and social research particularly, the data available for statistical analysis are often very irregularly distributed between various important classifications, either for administrative reasons or by the nature of the problem. Analysis of variance, multiple regression analysis, and similar statistical techniques then become exceedingly laborious; and I suspect that the results obtained are often too imprecise to justify the amount of work. I suggest that there is scope here for simpler and less time-consuming methods. I present below a method devised for one such situation; it is essentially an approximation to multiple regression by means of graphical analysis and χ^2 tests. Despite its extreme simplicity, I have never seen it used before. It is best explained by an example.

Example: Relation between Birth Weight and Pre-eclampsia

The records of all first single births in Aberdeen for the years 1948–53 have been examined. Births before the thirty-seventh week of gestation

and births for which the length of gestation could not be determined are not discussed here. Births in the thirty-seventh week and later totalled 4 323 and included 82 births to mothers who developed moderate or severe degrees of pre-eclampsia. We wanted to determine whether birth weight is influenced by the occurrence of pre-eclampsia.

Birth weight is known to vary with the length of gestation, the sex of the child, and the general size of the mother; we used maternal height as an indication of maternal size, since indices like weight or weight-height ratio change markedly during pregnancy. We knew from previous investigations that the relation between birth weight and gestation period is non-linear, while that between birth weight and maternal height is nearly linear. Sex differences in birth weight tend to increase as the length of gestation increases. In any attempt to assess the effect of pre-eclampsia, these three variables ought to be taken into account.

The obvious method of analysis would be to use multiple regression and to compare mean birth weights of the normal and the pre-eclamptic births after adjustment for gestation period and maternal height. Separate regressions would be needed for the two sexes. The non-linearity of the regression would make the computational work very considerable. The pre-eclamptic group was small, having only thirty-nine male and forty-three female babies. Hence any numerical estimate of the difference in birth weight between the pre-eclamptic and the non-pre-eclamptic group would inevitably be subject to large error. Since we can do little but try to establish whether or not a consistent difference exists, it seems extravagant to undertake a laborious multiple regression analysis.

In fact such an analysis can be replaced by an exceedingly simple procedure. Before embarking on any analysis of this type, we usually know enough about the problem to assess which of the variables involved is the most important. Here there is no doubt that the regression on the length of gestation is of primary importance, and we therefore try to take account of this variable as fully as possible. After that, we allow for the remaining variables by subdividing the data into a number of groups that are broadly homogeneous with respect to these variables.

All records, including the pre-eclamptic series, were first divided according to sex of child, and then further divided according to whether the mother was tall, of medium height, or short. The height groups were chosen in such a way as to form three classes of nearly equal width and divided the total number of cases almost in the proportion 1 : 2 : 1. This produced six sub-groups differing markedly in average birth weight. Each sub-group was then classified according to length of gestation and the median birth weight found for each week. Fig. 1 shows these median birth weights plotted against the gestation period, for each of the sub-groups. Curves were then drawn to represent the trends of the medians. In this instance, as in many others, such curves could be drawn satisfactorily by eye, provided that some attention was

paid to the varying reliability of the medians and the trend of the curve not unduly disturbed by a point relating to a relatively small number of observations. If the number of observations was very large, similar curves could be drawn for the upper and lower quartiles and these could be used in the subsequent analysis.

Attention was now focused on the eighty-two pre-eclamptic births. The birth weight of each of these babies was plotted on its appropriate section of Fig. 1, so giving a series of points some of which were below and some above the curves. Had any points fallen exactly on the relevant curve they would have been neglected in the subsequent counts. From the curves, these points can be immediately classified as falling within the upper or within the lower fifty per cent of their appropriate birth weight distribution after allowing for sex, gestation period, and height. If there were no association between pre-eclampsia and birth weight, the points should occur on an average just as often above as below the curve. Hence a χ^2 or binomial distribution test can be used to test the significance of the deviation of the actual number of points from the theoretical distribution on the hypothesis of no association. Moreover, Fig. 1 will indicate whether the same pattern of deviations emerges for different sex- and different height-groups and at various weeks of gestation. Often in such an analysis a visual inspection would be sufficient, but statistical tests of interaction between various sub-groups can be made by means of χ^2 if required.

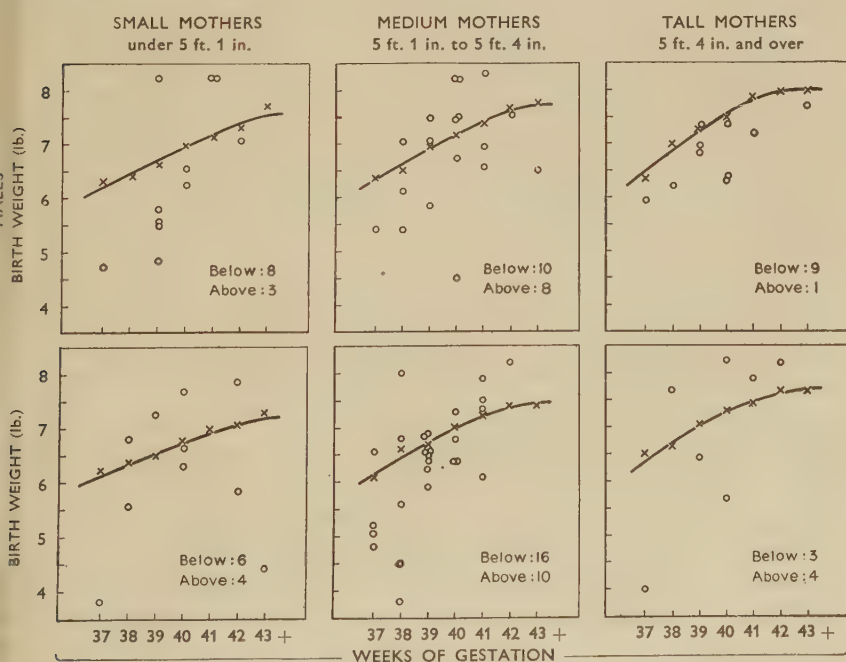


FIG. 1. Birth weight charts: first single births. The circles denote the weights in the pre-eclamptic group, and the crosses the median values before smoothing. The numbers of circles below and above the median line are given towards the bottom of the panels of the diagram.

The results of plotting are summarised in the three parts of Table I. Each of the parts shows a different aspect of the data.

TABLE I

Births to mothers who developed albuminuric pre-eclampsia: numbers of birth weights falling below or above median curves grouped according to (a) sex, (b) mother's height, and (c) length of gestation

	(a)		(b)			(c)		Totals
	Male	Female	Below 5 ft. 1 in.	5 ft. 1 in. to 5 ft. 4 in.	Above 5 ft. 4 in.	Up to 40 Weeks	41 Weeks and Over	
Below median	27	25	14	26	12	42	10	52
Above median	12	18	7	18	5	20	10	30
Totals . .	39	43	21	44	17	62	20	82

Of the eighty-two births, fifty-two fall below the median curves. A test using Yates's correction gives $\chi^2 = 5.38$. If there were no association between pre-eclampsia and birth weight, so large a value would arise by chance only in 2.5 per cent of tests, so that the data provide significant and fairly conclusive evidence that births to pre-eclamptic mothers tend to be below average weight. A glance at the table indicates that there is no evidence that the proportion of light-weight births differs in the two sexes or in the three maternal height groups. There is, however, some suggestion from the last part of the table that the inequality in numbers of light-weight and heavy-weight births disappears after forty weeks of gestation. A χ^2 calculated for the interaction of light and heavy with gestation period just reaches the ten per cent significance level, and so does not convincingly support the view that the situation changes at about forty weeks. However, the figures in this section of the table are sufficiently striking for further investigation to be desirable on other series of records. Possibly the more severe cases of pre-eclampsia tend to be delivered early. Another possibility is that the onset of the disease affects foetal growth only when it coincides with a particular phase in the development of the foetus. It is not for the statistician to speculate further along these lines; if the tendency should be confirmed on a larger series of records, some medical explanation would have to be sought.

Although the point does not arise in the example given above, a word of caution as to the acceptance of interaction effects on the basis of summary tables of the type of Table I may not be amiss, since it is possible to get apparent interaction effects. Suppose, for instance, that two of the criteria of classification are highly correlated and that there is an interaction between one of them and the phenomenon studied; then the summary table for the other criterion will also be suggestive of interaction even if none existed in reality. In such cases the apparent

and the true interaction effects can be sorted out by a rearrangement of summary tables along the obvious lines, since all the necessary information is contained in the charts.

Discussion

Various faults, limitations, and special merits of the method just illustrated will now be summarised.

i) The method does not give an estimate of the magnitude of the difference between the two groups. In this example, as in many others, knowledge of the numerical difference is of relatively little importance and the difference, even if estimated, would be subject to too great a standard error to be useful. Demonstration of the existence of a difference is a sufficient pointer for further research. Clearly the method could not be used in circumstances that demand careful numerical estimation, except perhaps in a preliminary survey of the data.

ii) The method can be used only if the group under special study and the control group with which it is to be compared differ very considerably in size, and if the whole body of data is sufficiently extensive for the charts to be drawn satisfactorily. Such a situation is quite common in medical and social research.

iii) As used in this example, the charts are based on all cases; the median lines were based on the pre-eclamptic group as well as upon the controls. This means some loss in sensitivity of analysis and a conservative bias in the tests of significance. However, the pre-eclamptic group is so small that the effect of its inclusion must have been negligible. Of course, it would have been possible, and perhaps theoretically preferable, to base the medians upon the 4241 births exclusive of the pre-eclamptic series. The advantage of using the complete series is that the same charts can be used again in similar analyses, for any other small groups of cases extracted from the whole series on the basis of one particular feature.

iv) The method takes simultaneous account of all observations whatever their classification in respect of sex, height, and gestation period, makes no assumption about homogeneity of variances in the sub-groups, and makes practically no assumption about the type of relationship between variables. The number of different factors involved in the regression, and the manner in which the observations in the pre-eclamptic and control series are distributed with respect to these factors, scarcely affect the analysis.

The preparation of charts such as those in Fig. 1 does not require any more time than the initial processing required for any other type of analysis. Once the charts are ready, the testing itself is a simple process involving only the plotting of a few points and counting. The arrangement of the records into small homogeneous groups provides a useful check on the consistency of results. In the analysis of the pre-eclampsia data I found the time taken by the method under discussion

to be less than one quarter of that required for the standard computations of a more refined regression analysis. When the charts have been prepared, any other sub-group of cases from the same records can be analysed and tested in a few hours only.

(vi) I attach special importance to the fact that at no point in this analysis does the investigator lose sight of his basic data. The scheme of sub-classification adopted, and the preparation of the charts, can bring out points that might easily be overlooked in a routine numerical analysis for which the first step was the computation of many sums of squares and sums of products. Even if a full regression analysis were eventually to be made, a preliminary rapid survey of the material in this manner might be a considerable safeguard against the danger of overlooking important but not obvious features of the data.

(vii) The graphs and summary tables obtained during the course of this method are easily understood by colleagues who may have little or no knowledge of standard statistical techniques and who will have to interpret the results in biological terms. It would indeed be a severe limitation on the statistician if he were never permitted to employ methods beyond the understanding of the non-mathematician. Nevertheless, the reader of a statistical report will feel much more confidence in the validity of an argument if he has some general understanding of how the analysis was performed. Perhaps even more important than the increase in confidence that arises when the reader does not have to take everything on trust is that this gives him an opportunity to draw attention to points which the statistician may have overlooked. A statistical analysis is of no use unless its conclusions can be interpreted in terms familiar to scientists working in the field in which it originated. The initial writing of this interpretation, and its understanding by the reader, are helped if critical appreciation of the technique of statistical analysis adopted is not limited to the professional statistician.

I am indebted to Dr D. J. Finney and Professor D. Baird and his staff for criticism and help.

MULTIVARIATE ANALYSIS*

JOHN WISHART

Statistical Laboratory, University of Cambridge

In this expository article Dr Wishart brings together a number of related topics that are usually scattered about text-books and papers. The section headings are: general distribution of estimates of variance and covariance; Bartlett's decomposition theorem; Hotelling's T^2 ; the discriminant function; Mahalanobis's D^2 ; and analysis of dispersion.

Introduction

A good deal has been written on the subject of multivariate analysis, and there are some text-books out of which you would be able to get everything I am going to describe if you were prepared to dig patiently for it. Yet the subject continues to be difficult for at least three reasons. One is that on the theoretical side the mathematics is of an advanced character; another is that on the practical side the computations are tedious; while a third is that it is so easy to get lost in a maze of problems, solutions to which have grown up independently, without realising that there is a certain unity of structure which, if understood, is likely to simplify and reduce the amount that has to be remembered. I propose to deal with selected parts of the subject only, and that in a particular way, in an attempt to portray the analysis as a generalisation of familiar univariate analysis, and to impose a certain degree of simplicity and unity of structure. A fourth difficulty which might have been mentioned is that of coping with the different notations and definitions used by various authors; I shall try to use throughout only one notation which should be generally acceptable. There must be scope for the application of some of these methods in new fields; in particular I am unaware of much work done in industrial statistics, and I should like to see them tried there.

As the name implies, the study is one of the appropriate methods of analysis when a single variate measure x is replaced by a number of measures in the variates $x_1, x_2, \dots, x_p$, p variates in all. A typical sample of size n will have n measures for each of the variates. Thus n men might have height, weight, etc., determined for each; n mass-produced articles might have a number of quality and other determinations made on each, and so on. Taken together, the population from which the sample is supposed drawn will be a multivariate one. It may be assumed normal (a basic assumption for most tests of significance), then not only will there be parameters μ_i and σ_i (or

*Paper read to the Industrial Applications Section of the Royal Statistical Society (London Group) on 2nd April 1954.

$\alpha_{ii} = \sigma_i^2$) for the mean and standard deviation (or variance) of each variate, i.e. for i taking all integer values from 1 to p , but also there will be $\frac{1}{2}p(p-1)$ additional parameters ρ_{ij} (or $\alpha_{ij} = \rho_{ij}\sigma_i\sigma_j$) for the correlation coefficients (or covariances) for all unequal values of i and j from 1, 2 to $p-1, p$. If the α_{ij} ($i \neq j$) were all zero, the variates would be independent and the analysis would break up into p univariate analyses, but in general this is not so. Sometimes univariate analyses are performed, and in addition correlation or regression coefficients are estimated from the sample, but this will not be considered as falling within the domain of multivariate analysis except in very special cases. Generally, we shall obtain from the sample (a) estimates of the p population means by the formulae

$$\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ik} \quad \text{for all } i = 1, 2, \dots, p$$

and (b) estimates of the $\frac{1}{2}p(p+1)$ population variances and covariances by the formulae

$$a_{ij} = \frac{1}{n-1} \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j) \quad \text{for all } i, j = 1, 2, \dots, p$$

When $i = j$ we are estimating a variance, otherwise a covariance. a_{ii} is usually called s_i^2 while a_{ij} is the quantity $s_i s_j r_{ij}$, r_{ij} being the sample estimate of the coefficient of correlation. Note that the convention of dividing by $n-1$ to secure unbiased estimates is used throughout.

For the most part I shall deal with significance tests which are generalisations of the univariate significance tests and are the main stock-in-trade of the practising statistician. It is therefore appropriate to say something at the outset about these tests, and about the basic mathematical analysis which leads to them. Let us suppose, then, that we have a normal population in the variate x , and a random sample of n members from it. A basic theorem, which incidentally is becoming easier to prove these days, is that the joint distribution of all the values of x , which is just the product of n normal distributions (since the x 's are independent), is the product of two univariate distributions: (1) a normal distribution of the mean $\bar{x}$ of the sample, which includes the parameters μ and σ ($= \sigma^2$) and also n , and (2) a distribution of the type known as χ^2 (or gamma) of the sample estimate a of the variance σ^2 , which includes the parameter σ^2 and also n . Distributions (1) and (2) are therefore independent. The t -test for comparing $\bar{x}$ with some hypothesis as to the value of μ , or for comparing two sample means, follows by combining (1) and (2) to get rid of the parameter σ^2 . Many tests use the χ^2 -distribution which is basic to (2), together with its degrees of freedom property. Distribution (2) also leads, by a comparison of two independent estimates of variance, to the well-known F or variance ratio test, which is used as an extension of the t -test when more than two samples are considered conjointly for agreement of their

means, and which is also used much more generally in all analysis of variance work.

General Distribution of Estimates of Variance and Covariance

It is fashionable in these days to date all multivariate analysis proper from the generalisation which was made in 1928 of the above mathematical propositions.¹ We consider this time the *set* of measures $x_1, \dots, x_p$ in place of the single x , and then the sample will consist of such sets, supposedly drawn at random from a multivariate normal population. The basic theorem is now that the joint distribution of all sets, which is the product of n multivariate normal distributions, can be transformed to be the product of two distributions: (1) a multivariate normal distribution of the p means $\bar{x}_i$ ($i = 1 \dots p$) of the sample, which includes the parameters μ_i and α_{ij} ($i, j = 1 \dots p$) and also n ; and (2) a multivariate distribution, of generalised χ^2 type, of $\frac{1}{2}p(p+1)$ sample estimates a_{ij} of the population variances and covariances, which includes the parameters α_{ij} and also n . It follows, therefore, that the distributions (1) and (2) are independent. In theory, therefore, it was apparent at an early stage that there were possibilities of producing various tests of significance which would generalise the univariate tests. For the most part I shall be speaking of two main classes of test, although they occur in the literature in various forms; the first will be concerned with comparing the sample set of means $\bar{x}_i$ with a population set μ_i adopted on some hypothesis— p means involved and a composite test is required; this class will also include the more practically useful test of comparing for agreement two sample sets of means, namely $\bar{x}_{i1}$ and $\bar{x}_{i2}$, $i = 1, 2 \dots p$, again a composite test; the second class will derive from the comparison of two independent distributions of the form (2), with the intention of having a test for the comparison of three or more sample sets of means, or more generally for use in some sort of generalised analysis of variance and covariance, or ‘analysis of dispersion’, as Rao has called it.²

The mathematical difficulties are formidable, as can be guessed if we try to write down the general form of distribution (2). In the univariate case write A for the inverse of α , i.e. α^{-1} , multiplied by $\frac{1}{2}(n-1)$. Then the form of (2) is proportional to

$$A^{\frac{1}{2}(n-1)} \alpha^{\frac{1}{2}(n-3)} \exp(-Aa) da$$

only a gamma function, involving n , is left out).

In the *bivariate* case we write A_{11} , A_{12} , and A_{22} for the inverses of α_{11} , α_{12} , and α_{22} , each multiplied by $\frac{1}{2}(n-1)$. But the word ‘inverse’ here has a special meaning. The set α_{ij} form what is called a matrix

$$(a_{ij}) = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{pmatrix}$$

whose determinant $|a_{ij}| = \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{vmatrix} = \alpha_{11}\alpha_{22} - \alpha_{12}^2$.

The inverse of a_{ij} is the i, j element of the inverse matrix to (a_{ij}) namely

$$(a_{ij})^{-1} = \begin{pmatrix} a_{22} - a_{12} \\ -a_{12} & a_{11} \end{pmatrix} \div |a_{ij}|$$

It will assist those who are not very familiar with matrices, or with this notation, if I say that to obtain A_{11} , A_{12} , and A_{22} we multiply $\frac{1}{2}(n-1)$ by

$$\frac{1}{\sigma_1^2(1-\rho^2)}, \quad -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}, \quad \text{and} \quad \frac{1}{\sigma_2^2(1-\rho^2)}$$

respectively. In this notation $|a_{ij}| = \sigma_1^2\sigma_2^2(1-\rho^2)$. For this (bivariate) case the form of (2) is proportional to

$$|A_{ij}|^{\frac{1}{2}(n-1)} |a_{ij}|^{\frac{1}{2}(n-4)} \exp \left(-\sum_{i,j=1}^2 A_{ij}a_{ij} \right) \prod_{i,j=1}^2 da_{ij}$$

in which i, j take the values 1 and 2. Single quantities have been replaced by 2×2 determinants, the exponential is a linear compound of three a 's, and the distribution is trivariate (a factor left out contains two Γ -functions in n and a power of π).

In the general (p -variate) case the required form is proportional to

$$|A_{ij}|^{\frac{1}{2}(n-1)} |a_{ij}|^{\frac{1}{2}(n-p-2)} \exp \left(-\sum_{i,j=1}^p A_{ij}a_{ij} \right) \prod_{i,j=1}^p da_{ij}$$

in which i, j take all values from 1 to p . The determinants are of order $p \times p$, and the distribution is $\frac{1}{2}p(p+1)$ -variate. An element A_{ij} is $\frac{1}{2}(n-1)$ multiplied by a^{ij} , this being a notation for the i, j element of the matrix inverse to that of the a_{ij} . The main thing to remember is that the population parameters occur in $|A_{ij}|$ and the sample estimates in $|a_{ij}|$.

Bartlett's Decomposition Theorem

The complexity of this result is only concealed, not removed, by expressing it in a modern elegant matrix notation. But in spite of this a number of useful results have been obtained from it, or in some cases more directly. First, there is the very interesting Decomposition Theorem, as it is called, due to M. S. Bartlett.³ This is to the effect that our general distribution (2) can be split into a number of factors for various constructed estimates, which are therefore independent. Thus we can show how multiple regression analysis, which can be regarded as part of multivariate analysis (I shall assume that it is familiar to you and need not be described), can itself be analysed into a chain of operations, beginning with the simple regression of dependent variate on x_1 , then going on to regression on the 'partial variate $x_{2.1}$ (i.e. that part of x_2 not directly predicted by regression on x_1) and so on until all variates have been brought to account. The factorisation, and the discussion Bartlett gave to it, did a great deal to

ationalise and simplify regression analysis from the angle of significance testing, particularly on the point that a degree of freedom is lost for each variate eliminated but that otherwise the distributions remain of the same standard form. On the computational side it led to the development of an easy stage-by-stage method of doing multiple regression which requires little more than the working out of determinants of order 2×2 as illustrated above (e.g. $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$) instead of having to go in for large-scale matrix inversion. I have had some success in expounding this method to a general non-mathematical class of student⁴; it has also been adapted recently to the most general case of orthogonal polynomial fitting.⁵

Hotelling's T^2

Let us now consider the composite test of agreement of the sample means $\bar{x}_i$ with population means μ_i ($i = 1 \dots p$). In the univariate case we may rewrite the 'Student' test in the present notation. Let $\bar{x} - \mu = d$, and let s^2 , the sample estimate of variance, $= a$. Then the square of 'Student's' t may be written

$$t^2 = na^{-1}d^2$$

the quantity t follows the familiar distribution for which tables are available; alternatively t^2 may be considered as a variance ratio with degrees of freedom 1 and $n - 1$, and the F -tables may be used. Hotelling⁶ deduced the generalisation of this to p variates. We write $\bar{x}_i - \mu_i = d_i$, $d_j = \bar{x}_j - \mu_j$ and we use a^{ij} for the i, j element of the matrix inverse to the matrix (a_{ij}) of the sample estimates of variances and covariances. We can then calculate

$$T^2 = n \sum_{i,j=1}^p a^{ij} d_i d_j$$

where the summation is over the $\frac{1}{2}p(p+1)$ values got by giving the values 1, 2, . . . p to both i and j . It will be seen that it is a direct generalisation of t^2 , for in the univariate case a^{-1} is the inverse or reciprocal of a and there is only one d . In the multivariate case we have what is known as a quadratic form in the d 's. T^2 can be calculated from the sample data when the values of μ_i , with which the $\bar{x}_i$ are to be compared, are laid down or assumed as hypothesis. As regards the distribution of T^2 , Hotelling's important result is that $T^2/(n-1)$ is distributed as the ratio of two independent sums of squares of normal variates having degrees of freedom p and $n-p$ respectively. It will perhaps be more familiar to you if I put this in its variance ratio form. The quantity

$$\frac{(n-p)T^2}{p(n-1)} \quad \text{or} \quad \frac{n(n-p)}{p(n-1)} \sum_{i,j=1}^p a^{ij} d_i d_j$$

is a variance ratio (or F) with degrees of freedom p and $n - p$. The chief difficulty in carrying out this test is the inversion of the matrix of the a_{ij} to get the a^{ij} . This operation has by now become mechanised on computing machines. The interpretation of the test should be clear to you. If the result is not significant then, on the basis of the observed characters taken jointly, the sample is compatible with belonging to the population with the given μ_i ; if the result is significant the reverse is the case.

The more practical case of comparing the sets of means provided by two independent samples of data in p variates is easily provided for. This is the analogue of the Fisher extension of 'Student's' t -test to the difference of two sample means. As could almost be guessed by anyone present, we write $d_i = \bar{x}_{i1} - \bar{x}_{i2}$, the subscripts 1 and 2 denoting corresponding means for samples 1 and 2, and we calculate a_{ij} from the combined sums of squares and products of deviations of the sample observations from the sample means. Thus

$$a_{ij} = \frac{1}{n_1 + n_2 - 2} \left[\sum_1^{n_1} (x_{i1} - \bar{x}_{i1})(x_{j1} - \bar{x}_{j1}) + \sum_1^{n_2} (x_{i2} - \bar{x}_{i2})(x_{j2} - \bar{x}_{j2}) \right]$$

where the sample sizes are n_1 and n_2 . It is the new matrix (a_{ij}) that we now have to invert to get the a^{ij} , and now, corresponding to the former formula for T^2 with both sides divided by n , we have in the present case

$$\left(\frac{1}{n_1} + \frac{1}{n_2} \right) T^2 = \sum_{i,j=1}^p a^{ij} d_i d_j$$

The same form of distribution holds for $T^2/(n_1 + n_2 - 2)$ as for the old $T^2/(n - 1)$, except for modification of the degrees of freedom. The quantity

$$\frac{(n_1 + n_2 - p - 1) T^2}{p(n_1 + n_2 - 2)} = \frac{n_1 n_2 (n_1 + n_2 - p - 1)}{p(n_1 + n_2)(n_1 + n_2 - 2)} \sum_{i,j=1}^p a^{ij} d_i d_j$$

is a variance ratio (or F) with degrees of freedom p and $n_1 + n_2 - p - 1$. Comparison with F -tables will show whether the samples are compatible with having come from a common population or not, at any rate so far as these p characters are concerned.

The Discriminant Function

There are two other problems and tests that have been developed since Hotelling's day, which will be described because they attempt to do different things, even though, as we shall see, the appropriate significance test (at least on the null hypothesis) turns out to be identical with Hotelling's T^2 -test and so was not, from this point of view, a new discovery.

The first is the use of Fisher's Discriminant Function.⁷ There is an important problem of what may be termed 'species identification' in

more than one field. We should like to be able to say, after measuring p characters on an individual, or (better) measuring a sample of n such individuals, that each individual probably belongs to one or other of two groups (more generally to one of possibly many groups). Thus the physical anthropologist may take character measurements on an unknown skull or group of skulls, and may have on record mean and covariance estimates for those characters for samples of known identification—e.g. a set of Bronze Age or of Iron Age measurements—and he would like to place the unknown. The plant ecologist may discover an unknown species and from measurements made on its members will wish to be able to say to what known group it probably belongs. Or he may suspect that species differently named are in reality identical. Let the variate measures be $x_1, x_2 \dots x_p$. Fisher sought for a linear function

$$y = l_1 x_1 + l_2 x_2 + \dots + l_p x_p$$

which would best discriminate between two groups. The problem is to determine the l coefficients. Before seeing how this is done let us suppose that we have two samples, of sizes n_1 and n_2 , to compare in p characters, and that coefficients $l_1, l_2 \dots l_p$ have been determined from the combined data. We then have a sample of n_1 values of y , not by combining the l 's with the values of x of the first sample. Let their mean be $\bar{y}_1$. Similarly we shall have a sample of n_2 values of y , with mean $\bar{y}_2$ (using the same l 's). By an ordinary analysis of variance let us break up the total sum of squares of the $(n_1 + n_2)$ values of y into a part between sample means and a part within samples. The analysis gives

	DF	SS	MS
Between samples:	1	$S_1 = \frac{n_1 n_2}{n_1 + n_2} (\bar{y}_1 - \bar{y}_2)^2$	M_1
Within samples:	$n_1 + n_2 - 2$	$S_2 = \sum_1^{n_1} (y - \bar{y}_1)^2 + \sum_1^{n_2} (y - \bar{y}_2)^2$	M_2

The criterion for 'best' which was adopted was to choose the values of $l_1, l_2 \dots l_p$ which maximise the ratio M_1/M_2 of the mean squares. The usual procedure for the determination of a maximum leads to p equations in p unknowns to determine the l 's (or rather their ratios, since there will be an arbitrary element present). These are

$$\begin{aligned} l_1 a_{11} + l_2 a_{12} + \dots + l_p a_{1p} &= d_1 \\ &\dots\dots\dots \\ l_1 a_{p1} + l_2 a_{p2} + \dots + l_p a_{pp} &= d_p \end{aligned}$$

where the a_{ij} are the composite covariance estimates as defined above in Hotelling's T^2 (second case), and $d_i = \bar{x}_{i1} - \bar{x}_{i2}$, ($i = 1, 2 \dots p$). Solution of these equations will determine $l_1 \dots l_p$, thus determining

the discriminant function. It then follows that the maximised M_1/M_2 is

$$\frac{n_1 n_2}{n_1 + n_2} \frac{(l_1 d_1 + \dots + l_p d_p)^2}{\sum_{i,j=1}^p a_{ij} l_i l_j} = \frac{n_1 n_2}{n_1 + n_2} (l_1 d_1 + \dots + l_p d_p) \\ = \frac{n_1 n_2}{n_1 + n_2} \sum_{i,j=1}^p a^{ij} d_i d_j = T^2$$

The test of significance for the discrimination will be a test of whether the two samples could have come from the same population. Now we have the ratio of mean squares M_1/M_2 , and the appropriate test is one for the size of this ratio. But this is T^2 , and so the test of significance for discrimination is just the test for T^2 already given.

The difference consists in the problem which is being tackled. A linear function of the x 's denotes a direction in the p -space of the variates, and there is a problem of estimation to determine this direction. The significance test determines whether this direction coincides with a given (or estimated) direction for some race or species. Since it is useful in such problems to have the discriminant function calculated, we can now vary the method of calculation of T^2 as given in the previous section. We may solve the above linear equations first to find the l 's, and then find T^2 from $n_1 n_2 \Sigma(l_i d_i)/(n_1 + n_2)$. The solution of the equations, by the general method, involves finding the a^{ij} from the a_{ij} , i.e. inverting the matrix of the a_{ij} , so that about the same amount of work is involved as before.

I have chosen to illustrate the discriminant function problem by the comparison of two samples of data. This corresponds to Fisher's example of data of sepal length and width, and petal length and width, in samples of 50 observations of two species of iris. It therefore corresponds to the second of the two cases of Hotelling's T^2 . But there is obviously also a case corresponding to the first T^2 one. This would arise where a single sample was compared with a known population, i.e. population parameters $\lambda_1, \lambda_2, \dots, \lambda_p$ are given for a certain discriminant, and we wish to see whether the set $l_1, l_2, \dots, l_p$ calculated from a sample agree, or do not agree, with the population values. Various examples of the use of the discriminant function will be found in the literature. Moroney has one in his book⁸ but, if I may say so without offence, it is a somewhat degenerate example. He considers the marks obtained in three separate tests by two different groups of students, but of the three mean differences in marks, d_1, d_2 , and d_3 , d_2 and d_3 are both zero.

A number of supplementary tests of this character are described by Rao in his book.² A biologist might choose an arbitrary linear function of his own by combining the measures $x_1, x_2, \dots, x_p$ in some simple fashion, as by choosing the l 's to be +1 in some cases and -1 in others. One can then work out the extent to which the real discriminant function is an improvement over this empirical procedure. A small

amount of additional calculation will determine a T'^2 involving the empirical l_i . Then a statistic easily calculated from the ratio of T^2 to T'^2 proves also to be distributed as a variance ratio, and so can readily be tested. Rao also shows that if a discriminant function be calculated for less than the full number of characters available, one can test whether these are enough for discrimination by calculating a statistic from the ratio of the complete to the partial T^2 . This is also distributed as a variance ratio. A third extension mentioned by Rao relates to having four samples of data. A discriminant function calculated for the first pair will discriminate between two populations, A and B say. Similarly we can discriminate between populations C and D from the data of the second pair. Then, under certain assumptions, e.g. that the variances and covariances are equal for all four populations, we can test whether the two discriminant functions are significantly different by a somewhat similar test-criterion to the above, also following the variance ratio distribution.

Mahalanobis's D^2

Another line of development in the history of statistics started with Karl Pearson,⁹ who invented a coefficient of racial likeness (CRL) to assist in the identification and classification problems of the physical anthropologist. The idea behind this is a very simple one. If, in a univariate normal sample, $\bar{x} - \mu = d$, then nd^2/σ^2 is a χ^2 with one degree of freedom, and so for k independent d 's $\sum_{i=1}^k (n_i d_i^2 / \sigma_i^2)$ will be a χ^2 with k degrees of freedom. For the difference between two independent samples we can read d as $\bar{x}_1 - \bar{x}_2$ and σ^2/n as $\sigma_1^2/n_1 + \sigma_2^2/n_2$. Pearson applied this to the p character-measures of a multivariate sample, ignored the fact that these were not independent, and inserted sample estimates for σ_1^2 and σ_2^2 . His coefficient was then proportional to

$$\sum_{i=1}^p \left\{ \frac{(\bar{x}_{i1} - \bar{x}_{i2})^2}{s_{i1}^2/n_1 + s_{i2}^2/n_2} \right\}$$

and he provided a standard error for it. It is therefore a crude measure of the 'distance' between two samples, in which the x 's were usually skeletal measurements. The object was to see how alike or unlike two species might be on the evidence of the remains. Various modifications were made to this coefficient from time to time, and in the thirties it became obvious to Mahalanobis and his co-workers in India that what was needed was to replace an expression like

$$\sum_{i=1}^p (a_i^{-1} d_i^2)$$

by the notation we have used in discussing Hotelling's T^2 by the quantity

$$D^2 = \sum_{i,j=1}^p a^{ij} d_i d_j$$

in which a^{ij} and d_i, d_j have exactly the definitions accorded to them above in the second of the two Hotelling T^2 cases. This actual form of D^2 was only introduced within recent years by Rao,² but in the form pD^2 the coefficient was fully considered by Mahalanobis in 1936¹⁰ (although sample numbers, and not degrees of freedom, were used in calculating the a_{ij}). Three actual forms were defined. In Rao's notation these are:

(1) Population Distance

$$\Delta^2 = \sum_{i,j=1}^p \alpha^{ij} \delta_i \delta_j$$

where α_{ij} are the population variances and covariances, assumed equal for two populations, and $\delta_i = \mu_{i1} - \mu_{i2}$

(2) Sample Distance with Population Dispersion Known

$$D^2 = \sum_{i,j=1}^p \alpha^{ij} d_i d_j$$

in which we compare two samples, and $d_i = \bar{x}_{i1} - \bar{x}_{i2}$, but are given the α_{ij} .

(3) Studentised Sample Distance

$$D^2 = \sum_{i,j=1}^p a^{ij} d_i d_j$$

Geometrically, D^2 is the square of the length of a vector in the p -space, so that in this problem we are concerned with the length of the vector as opposed to its direction in the discriminant problem. But for a test of significance, since

$$D^2 = \sum_{i,j=1}^p a^{ij} d_i d_j = \frac{n_1 + n_2}{n_1 n_2} T^2,$$

the distribution of D^2 in the null case, i.e. when $\Delta^2 = 0$, as worked out by Bose and Roy in 1938,¹¹ is of course the same as the distribution of T^2 worked out by Hotelling in 1931. However, Bose and Roy broke new ground by deriving at the same time the 'non-central' distribution of D^2 , i.e. the case $\Delta^2 \neq 0$.

It is usually the null hypothesis which is assumed in practice, and so the variance ratio (F) test given at the end of the section on Hotelling's T^2 is the one to apply. There are a number of examples of the test as applied by the physical anthropologists. It was carried out systematically as part of a large anthropological survey in India, and Rao has used D^2 (or D) as a measure with which to classify races and tribes. The CRL has been superseded by D^2 at the Duckworth Laboratory, Cambridge. Mulhall and Jones have made an important study of the data of African tribes, and Miss Fereday has calculated D^2 between her own British neolithic data and the published data of other more or less contemporary racial stocks, in order to trace affinities and to classify generally. The calculations are laborious, and this has led to approxi-

ations which depart from strict mathematical orthodoxy, although this probably does not make much difference in practice. There were many races to compare in the Indian survey that an average 'stock' matrix was determined once and for all, and used for each pair of samples to be compared. Miss Fereday's sample was so small that the matrix was 'borrowed' from a known long series which was certainly not British neolithic. This is equivalent to using the 'unstudentised' form of D^2 (see (2) above), in which case the distribution would be approximate to χ^2 form with p degrees of freedom.

The supplementary tests given at the end of the section on the discriminant function were described by Rao in terms of D^2 , not T^2 , but the conversion is easily made by noting from above that T^2 is just $n_1 n_2 D^2 / (n_1 + n_2)$. As they are all null tests of significance they may be thought of as discriminant or distance problems at will, except where an actual discriminant function has to be estimated.

Analysis of Dispersion

The name 'analysis of dispersion' was suggested by Rao for the multivariate equivalent of analysis of variance in univariate problems. When we pass from two samples to the consideration of three or more, or in general to the data of some experiment set up in factorial or other form, we have a need for tests additional to the t -test for the comparison of two means. An important class of such problems is dealt with in the analysis of variance procedure. Thus with k independent samples we can test for homogeneity, i.e. for equality of the k population means, by dividing the total sum of squares into a part between sample means and a part within samples, the test criterion being the variance ratio F (or more strictly the mean-square ratio). In more complicated cases certain sums of squares are first eliminated, and also individual components of the 'treatments' sum of squares are isolated. In multivariate analysis the first thing to note is that a sum of squares in one variable

$$S = \sum_i^n (x_i - \bar{x})^2$$

'blown up' in p variates into a matrix of sums of squares and products S_{ij} , $i, j = 1, 2, \dots, p$, where

$$S_{ij} = \sum_i^n (x_i - \bar{x}_i)(x_j - \bar{x}_j)$$

when $i = j$ we have a sum of squares, when $i \neq j$ a sum of products. The S_{ij} in this simple case will have $n - 1$ degrees of freedom, and division by $n - 1$ yields the a_{ij} already defined (for a single sample). Now consider for simplicity the analysis of dispersion appropriate to data given in k independent samples of sizes $n_1, n_2, \dots, n_k$ for each of p variates. The total $n = \sum_{i=1}^k n_i$ observations will yield a total S_{ij} defined as above, with $n - 1$ degrees of freedom. For all values of i and j this

will give us a 'total' dispersion matrix (S_{ij}). The element S_{ij} can now be split up by elementary methods into the two parts: (1) a sum of squares or products between sample means, which we shall call Q_{ij} in conformity with Rao's notation, having $k - 1$ degrees of freedom; for all i and j this will give us the 'between' dispersion matrix (Q_{ij}); and (2) the combined sums of squares and products within samples, which we call W_{ij} , leading to the 'within' dispersion matrix (W_{ij}) with $n - k$ degrees of freedom. Our analysis of dispersion then takes the simple form

			DF	SP matrix
Between	..	..	$k - 1 = q$	Q
Within	..	..	$n - k = w$	W
Total	..	..	$n - 1 = q + w$	S

if we use Q for (Q_{ij}) , etc. Note that $W_{ij}/(n - k)$ is a generalisation of the a_{ij} defined earlier when comparing two samples; for there will be k sums of products in the numerator and k degree of freedom components in the denominator.

In a more general case there will be more than two components in the analysis of dispersion. But one, analogous to the above 'between' will be a 'treatment' component, with the dispersion matrix Q measuring a deviation from some hypothesis (in the special case above this is equality of population means in all characters) and having in general q degrees of freedom, say; the other, analogous to 'within', will be an 'error' component measuring dispersion in data conforming to hypothesis and having w , say, degrees of freedom. These two components isolated from the rest of the analysis and then summed, will give the equivalent of the above table, the sum S having $q + w$ degrees of freedom. This is shown in the table in the alternative presentation of degrees of freedom.

For a test of the hypothesis (equality of means, or something more general) Wilks¹² developed a Λ (lambda) criterion. The principle will be readily grasped by analogy with univariate methods. On the null hypothesis the two 'mean product' matrices developed from Q and W will both estimate a single population dispersion matrix, while departure from hypothesis will change Q relative to W . The distribution problem has not been completely solved in full generality. Nor is it a generalisation of the F -test, but rather of its 'Beta' analogue. [Some of you will know that the variance ratio distribution of $F = v_1/v_2$, where $v_1 = S_1/n_1$ and $v_2 = S_2/n_2$, is equivalent to a 'Beta' distribution of $S_1/(S_1 + S_2)$]. Wilks's criterion is

$$\Lambda = |W| \div |S|$$

where the numerator is the determinant of the 'within' or 'error' matrix (W_{ij}) and the denominator is the determinant of the total matrix (S_{ij}).

The special cases in which the distribution of Λ has been found can be put in variance ratio form, and are listed below:

) For Any Number of Variates

	Variance ratio	Degrees of freedom
$q = 1 \quad \dots$	$\frac{1 + w - p}{p} \cdot \frac{1 - \Lambda}{\Lambda}$	p and $1 + w - p$
$q = 2 \quad \dots$	$\frac{1 + w - p}{p} \cdot \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}}$	$2p$ and $2(1 + w - p)$

) For q greater than 2

	Variance ratio	Degrees of freedom
$p = 1 \quad \dots$	$\frac{w}{q} \cdot \frac{1 - \Lambda}{\Lambda}$	q and w
$p = 2 \quad \dots$	$\frac{w - 1}{q} \cdot \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}}$	$2q$ and $2(w - 1)$

These formulae would enable us, for example, to test exactly for equality of means from the data of three samples in any number of variates, or to test any number of bivariate samples. To go further we must have recourse to M. S. Bartlett's ingenious χ^2 approximation.¹³ We take the natural logarithm of Λ (a negative number) and multiply by $-\frac{1}{2}(2w + q - p - 1)$. If we call this (positive) quantity V , the approximate result is that V may be taken to be a χ^2 with pq degrees of freedom. For reasonably large $m = \frac{1}{2}(2w + q - p - 1)$ this is quite accurate, and significance points may be obtained from the χ^2 table. For smaller m it should be noted that exact probabilities corresponding to given V may be got by utilising Rao's expansion,² though it is not quite so easy to find a V inversely for a given significance point.

This is not the end of the story, for it may sometimes be advantageous to split up Λ into a product of independent factors (or, what comes to the same thing, to decompose Bartlett's χ^2 into additive components). Again, it may be desirable to eliminate an underlying variate or variates from all the p variates under action (as in covariance analysis of one variate) before applying the Wilks criterion to the residual Q and W dispersion matrices. But this seems to be a suitable point at which to stop in giving a brief introductory account of multivariate analysis, with the final remark that there are other aspects of the subject, such as canonical correlation study and factor analysis, that have not even been mentioned. What is needed is for some of the material, at least, of the methods expounded to become a part of the general stock-in-trade of the practising statistician, and for their field of application to be widened to all cases in which several variates can be measured and could, with advantage, be considered conjointly.

REFERENCES

- ¹ WISHART, J. (1928). 'The generalised product-moment distribution in sample from a normal multivariate population', *Biometrika*, **20A**, 32.
- ² RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. Wiley and Sons, New York; Chapman and Hall, London.
- ³ BARTLETT, M. S. (1933). 'On the theory of statistical regression', *Proc. Roy. Soc. Edinb.*, **53**, 260.
- ⁴ WISHART, J. (1950). *Field Trials II: The Analysis of Covariance*. Technical Communication No. 15, Commonwealth Bureau of Plant Breeding and Genetics, Cambridge.
- ⁵ WISHART, J., and METAKIDES, TH. (1953). 'Orthogonal polynomial fitting', *Biometrika*, **40**, 361.
- ⁶ HOTELLING, H. (1931). 'The generalisation of "Student's" ratio', *Ann. Math. Statist.*, **2**, 360.
- ⁷ FISHER, R. A. (1936). 'The use of multiple measurements in taxonomic problems', *Ann. Eugen. Lond.*, **7**, 179.
- ⁸ MORONEY, M. J. (1953). *Facts from Figures* (second edition). Penguin Books, London.
- ⁹ TILDESLEY, M. L. (1921). 'A first study of the Burmese skull', *Biometrika*, **13**, 176.
- ¹⁰ MAHALANOBIS, P. C. (1936). 'On the generalised distance in statistics', *Proc. Nat. Inst. Sci. Ind.*, **11**, 49.
- ¹¹ BOSE, R. C., and ROY, S. N. (1938). 'The distribution of the studentised *D* statistic', *Sankhyā*, **4**, 19.
- ¹² WILKS, S. S. (1932). 'Certain generalisations in the analysis of variance', *Biometrika*, **24**, 471.
- ¹³ BARTLETT, M. S. (1930). 'Further aspects of the theory of multiple regression', *Proc. Camb. Phil. Soc.*, **34**, 33.

LETTER TO THE EDITOR

Theoretical and Practical Workers in Statistics

THE modern theory of statistics, especially as applied to industrial research, stems largely from the work of people who were originally concerned to find new methods of tackling practical problems that they or their colleagues in research were faced with. The success of the founders of modern statistical theory needs no stressing in the pages of this journal. Its measure can be judged by the diverse fields in which the theory is practised, many applications being quite different from those in which the theory was originally developed; the number of experiments designed and analysed statistically is probably greater today in industrial than in agricultural research.

One result of this development is that there is now a well-defined group of people who are applying statistical methods as their profession; their everyday tools are the theories which were novel fifty and thirty years ago, and their number is increasing. There is also a considerable amount of theoretical work in statistics being done parallel to the practical applications, but nowadays this is largely confined to the universities, with some notable exceptions. The situation appears to be developing, therefore, that whereas theory and practice originally stemmed from the same person, today there are two distinct groups performing these two functions separately. This in itself is, of course, as it should be and is a sign of maturity of the subject. It does, however, have important consequences for workers on both the theoretical and practical sides.

For the practising statistician it means that there is a growing problem to keep up to date with new developments in theory. This problem is accentuated by the fact that his working time may be fully occupied in applying well-established techniques, and these techniques enable him to solve the large majority of his problems. Nevertheless an appreciation of current developments in theory is a desirable stimulus for any group of professional workers, and particularly so in statistics.

There also arises the situation where the practical statistician is confronted with a problem which is not amenable to the better-known techniques, or he feels that a better answer might be obtained by a more refined method. If such a method exists it may involve a lengthy search through the literature, and an even more lengthy process of understanding the appropriate paper. He may also have to decide whether in fact the effort of learning and applying the new method is worth while. If a partial answer can be given by well-known techniques relatively quickly, the time saved may be more usefully spent on other work in hand, which he knows he can tackle with every hope of success. The applied statistician is faced with an even more difficult situation if he meets a problem to which there is no suitable theoretical solution. In this case he may try to do some theory himself, if he has the ability, and again he must decide if the time he spends on it is justified.

There may also be an effect on the theoretical work produced if there is a great division between theory and practice. As theoretical statistics

becomes more and more complex mathematically, there is a danger that it may be directed to the solution of mathematical rather than realistic statistical problems. I appreciate that a function of theoretical research is to anticipate application as well as to solve existing problems, but I cannot help feeling that an increasing amount of published work in theoretical statistics is focused more on the post-graduate examiner than on eventual application.

If this is a fair assessment of the position there may be a need for establishing closer contact between theoretical and practical statisticians. I am aware that some means already exist whereby contact can be made—for example through meetings of the various sections of the Royal Statistical Society through the Questions and Answers pages in this and other journals; at conferences; and through personal correspondence. Nevertheless, I feel that more needs to be done.

It may be useful to have short courses or summer schools run by the universities on particular new topics in mathematical statistics. These courses need only contain sufficient detail of technique to enable the student to get a grasp of the scope and power of the methods being taught, so that he may be able to judge whether a practical problem encountered during his work could be tackled by a new approach. The time for learning the details of the technique would be just before it was applied to the actual practical problem. Courses of this type might also provide an opportunity for practising statisticians to put some of their own unsolved problems to their academic colleagues.

A further help might be some sort of selective review of recent developments in theory, on the lines of Dr Irwin's articles that appeared in the *Journal of the Royal Statistical Society* during the 1930's. Such a review need not include a mention of every paper published, nor need it be a series of abstracts. More usefully, it would be the opinions of a competent writer on current trends in research in mathematical statistics, particularly in the branches that may have useful applications, with a little expository material included. The review could best serve its purpose if it was not written in the impartial style of orthodox abstracting, but reflected clearly the authors' opinions as to what was important and what was not. I suggest, sir, that the pages of *Applied Statistics* might be a suitable place for such articles, as a supplement to the expository papers that appear from time to time.

I would like to emphasise that the problem of communication between theoretical and practical statisticians I am considering is different from the similar one that exists between statisticians and non-statisticians. Much has been written and said on the topic of putting over statistical ideas to the layman, and it is part of almost every practising statistician's job to explain his methods and conclusions to the engineer, scientist, or industrialist. I am not concerned with this problem here. I am thinking only of the passage of information, in both directions, between statisticians who have a grasp of the elementary principles of the subject and are applying these principles in their daily work on the one hand, and the innovators of the new techniques and the new lines of thought on the other.

The preservation of the link between theory and practice is a problem not unique to statistics, and one hears many other scientific workers concerned with the same difficulty. Nevertheless, I believe that there are certain peculiar features which make it a particularly important problem

statistics. In my opinion statistics is not a branch of pure mathematics, or even part of applied mathematics, and theory and practice should reflect this view.

A. GLASKIN

17th February 1955

[We welcome articles of the kind mentioned by Mr Glaskin.—EDITORS]

NOTES AND COMMENTS

THE Japanese journal *Statistical Quality Control* gives evidence of considerable activity in the field of quality control in Japan, since the 1954 volume contains over 900 pages. It is printed in Japanese characters and so is unintelligible to most Western readers—although mathematics is presented in Western symbols and Western proper names are printed in roman type—but English abstracts of some of the papers indicate that they cover the usual topics of the subject. The cover is elaborately designed with numerous small panels, each containing a mathematical or statistical device (a curve or formula used as a design), or a small symbolical representation of an industrial process.

★ ★ ★

A tutorial course of weekly lectures which has been held at Management House,* London, this winter is the result of an expression of interest by several leading firms in having staff trained in the use of statistical methods. The course was attended by sixteen technically qualified men in executive or research positions in the eight firms represented. It is notable that these men were nominated by their firms and that the decision to attend was made after some initial experience of the application of the methods to the work of the firm. There must be many other firms to whom the methods would be equally valuable.

The purpose was to provide an opportunity for the technologist, already a specialist in one subject, to acquire these techniques of dealing with the variability commonly encountered in production processes or in the results of experiments. The lectures aimed at giving him an acquaintance with the working tools of the statistician up to and including the basic designs for an efficient experiment.

The lectures came to an end in early April and are to be supplemented by a short residential course at Cambridge in the early summer on practical statistical calculations. This should serve to consolidate the knowledge already gained and to provide opportunities for discussion of difficulties. Being residential, this supplementary course will appeal to some unable to attend the London lectures; provision is therefore being made for a modified course to run parallel to the main course with a shortened syllabus suitable for those beginning their study of the subject.

It is hoped to repeat the courses next session and to start the weekly lectures in October. For information write to Mr E. D. van Rest, 11 Nightingale Road, Hampton, Middlesex.

★ ★ ★

In a review in the *Times Literary Supplement* of 11th March 1955 appeared the statement:

‘The method of sampling has been justified by its success in prediction and statesmen will rely on it more and more for interpreting the oracle of

* Although the lectures were held at Management House they were not organised by the British Institute of Management.

mocracy. No doubt it is a mere reactionary prejudice to regret that men and women should be so much of a pattern that what is true of a "random sample" of one in 100 (or 1 000) is true of all.'

This seems to disclose a lack of sympathy with the sampling method coupled with a grudging acceptance of it that is perhaps part of the pattern attitudes of people to the statistical approach in general. True it is that sample patterns can be discerned in the behaviour of people—one has only go to a football match or listen to a radio audience reacting to a radio mic to realise how simple some of these patterns can be. Statistical method gives them numerical description, and only relatively simple patterns can be estimated from a sample—obviously a sample of a few hundred people can tell little of the idiosyncrasies of many thousands. In dealing with the patterns we—statisticians, social scientists, and people of that sort—may sometimes overlook the idiosyncrasies, and perhaps this is what the writer we have quoted really regrets; to pay attention to the patterns is not to deny the existence of the idiosyncrasies. The statistician would certainly think it reactionary to regret that there is value in paying attention to the patterns.

★ ★ ★

The Association of Incorporated Statisticians has recently revised the syllabus for the examination for Registered Statistical Assistants. The main change is that an opportunity is now provided for specialisation in economic, industrial, and medical statistics. There is a three-hour paper in elementary statistical methods, in presentation and computation, and in the sources and methods of collection or analysis relevant to the special fields. The minimum age for candidates is 18. Copies of the syllabus and specimen papers can be obtained from the Secretary, Association of Incorporated Statisticians, Park Lane, London, W.1.

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

List and Reports of Meetings

BIRMINGHAM AND DISTRICT GROUP

- 1st December 1954. Simpler Design of Experiment: R. C. Curnow
5th January 1955. Plant Loading and Utilisation in Job Production
Stafford Beer
See the article in *Applied Statistics*, 4, 1.
2nd February 1955. Easy Profitable Applications of Statistics in Industry
M. J. Moroney

GLASGOW GROUP

- 13th January 1955. Operations Research applied to Materials Handling
J. Murdoch
8th February 1955. Design of Experiments to secure Process Improvement
G. E. P. Box

LEICESTER GROUP

- 17th February 1955. Forecasting Sales: Miss J. Keen

LONDON GROUP

- 26th November 1954. The Use of Statistics in Personnel Management
A one-day conference on the above subject was held at the Connaught Rooms, Great Queen Street, London. It was organised jointly by the British Institute of Management, the Institute of Personnel Management and the Study and Industrial Applications Sections (London Group) of the Royal Statistical Society.

The object of holding such a conference was to bring the statistical and personnel specialists to a common forum, whereby they could define the overlapping spheres of interest in order that a better understanding might result of the problems to be solved and of the methods available for solving them. Over 180 delegates attended, of whom approximately two-thirds were drawn mainly from the management or personnel fields, the remainder being primarily statisticians.

Mr John Ryan opened the proceedings, and the following papers were given: 'The Scope and Problems of Personnel Management' by H. E. Roth; 'Getting the Most out of Facts and Figures' by R. L. Webster; 'Making sure that the Methods are Sound' by M. G. Kendall and P. E. Vernon; and 'Finding out what People Think' by H. J. Eysenck. The day's proceedings were summed up by Mr Bernard Ungerson.

rd December 1954. Some Fundamental Problems in the Subjective Appraisal of Foodstuffs: R. Harper

The functions served by the subjective appraisal of foodstuffs were examined. The judgements concerned seem to fall into two distinct types: the grading type and the analytical type. The analytical type is the special province of the laboratory panel, and it is against these judgements that existing instrumental tests must be evaluated. Human judgements are affected by many complex influences which do not readily fit in to investigations based on the assumptions of factorial designs. Such influences were examined, with special reference to those instances in which the 'qualities' assessed are multidimensional. By way of example certain studies on Cheshire cheese, involving the use of factor analysis, were outlined and criticised. Finally Dr Harper appealed for the development of statistical techniques that conform to the demonstrated 'properties' of the data collected.

th February 1955. A Multi-head Sorting Machine for the Grading of Small Ceramic Components: A. N. James and D. Boswell

Mr Boswell began by describing the small ceramic spacers used for separating the electrodes in cathode-ray tubes and stressed the importance of the lengths of individual spacers being correct within plus or minus 0.001 in. in order to ensure that the electrodes were parallel. He described the design and operation of a machine that had been specially devised for the purpose of sorting these small spacers according to their deviation from a predetermined target, and a demonstration was given of the apparatus in operation.

Mr James then described how various statistical tests had been carried out to determine the efficiency of operation of the machine and explained the difficulties encountered due to overlapping of groups and central bias in the machine's operation. He went on to describe sequential and control chart tests which could be applied to determine whether the machine was sorting efficiently. The results of some of these tests had resulted in modifications to the design of the machine, and at all stages there had been close liaison between the designers and the statisticians.

MERSEYSIDE GROUP

th January 1955. Probability in Industry: Miss E. A. G. Knowles

Dr Knowles discussed briefly von Mises's theory of probability and her objections to it. After indicating certain modifications which would make the theory more acceptable to the practising statistician, she discussed a practical example showing how probability theory had been applied to a quality control problem.

th February 1955. Some Experiences in the Economic Assessment of New Industrial Methods: K. P. Norris

Mr Norris discussed the situations that occur when a choice must be made between two or more processes. Various criteria can be used, but the decision will ultimately be based on costing figures. The various difficulties encountered in estimating costs were examined and practical examples quoted. The concept of 'rate of depreciation' was criticised and an alternative criterion proposed which was more realistic and more widely applicable. The speaker ended with a discussion of the effects of taxation on the development of new processes.

NORTH-EASTERN GROUP

8th December 1954. The Advertisers' Use of Statistics: P. Harper

19th January 1955. The Design of Industrial Experiments: N. L. Frankl

SHEFFIELD GROUP

25th November 1954. A Mathematical Treatment of Reasons for Failure
K. F. Lane

The problem considered was one where an item and component may fail in service for more than one reason. Such an item is an ingot mould, which can fail because of a crack or because of a worn interior surface. A theory was proposed to explain the behaviour. The initial assumptions were described, together with mathematical formulae that had been devised from them. The application of the theory to the analysis of an example was described, together with the results that determine the patterns which should be present if the theory holds. Actual figures for ingot moulds were quoted, and it was found that they did exhibit these patterns.

27th January 1955. Inventory Control in Firms: K. S. Lomax

The economic processes of the firm were described and the advantages and disadvantages of large and small inventories were discussed. The factors which influence inventory size were considered and the following four points were made: (i) that few of these factors are under the direct control of the firm; (ii) that the relationships between the factors and the optimum size of inventory are vague and hard to express; (iii) that most constraints are inequalities rather than continuous functions; (iv) that at least one factor, namely demand, varies randomly.

The various ways of calculating expected net revenue were described, in cases where the level of inventory is controlled by the firm and the distribution of demand is known. Systems of inventory control in this country were contrasted with those in use in the USA. A simple example was considered in which the optimum inventory level had been arrived at by two methods. The first, quoted from Fry, aimed only at minimising loss of sales. This was contrasted with the proposed method, where a different result was obtained when a price unit was assumed and storage costs taken into account.

24th February 1955. Smoking and Lung Cancer: R. Doll

The various methods of investigation were discussed and answers given to some of the criticisms levelled at the methodology used in past investigations. Details were given of the major retrospective investigation 'Aetiology of Carcinoma of the Lung', the results of which were published in 1952. A current prospective investigation was discussed in which an attempt is being made to associate the smoking habits of doctors with their mortality rates.

SOUTH WALES GROUP

2nd December 1954. A Practical Approach to making Statistical Decisions
G. A. Barnard

After describing the literature on decision theory Professor Barnard showed how making a wrong decision would involve a financial loss which could be estimated. Decision functions or rules could therefore be compared

on the basis of the average loss incurred in the long run and enabled one to define a class of decision rules which could be called good rules. The class of Bayes solutions was in fact a complete class of good decision rules, but the appropriate solution depended on a knowledge of what was expected from the process in the long run. Without such knowledge Wald proposed the minimax principle, which chose the solution that minimised the losses under the worst possible situation, but this cannot be recommended as a good principle. Professor Barnard also considered Bayes's principle to be bad.

He then considered the problem of sampling inspection by attributes. When the process curve consisted entirely of two distinct qualities, say acceptable and rejectable quality, in known proportions, then the decision rule was simply determined. A means of approximately representing any process curves in this way was demonstrated and shown to give a similar decision rule to that based on the full process curve. However, this approximation did not work for the problem of sampling by variables because the amount of information obtained from the sample itself in this latter case was so very much greater.

14th January 1955. Sensory Tests in Industrial Research: J. R. Williams

This lecture is substantially the same as that given to the Merseyside Group on 1st December 1953 and reported in *Applied Statistics*, 3, 56.

15th February 1955. Does Normality Matter?: G. E. P. Box

Dr Box considered the assumptions made in standard statistical tests and classified the tests according to whether the assumptions affected the test to any marked degree or not. In a comparison of means by an analysis of variance the assumption of normality of variance was unimportant, and that equality of variances was not important when all the means were determined with an equal number of observations; with very unequal numbers of observations, however, the latter assumption was important. In experiments on growth curves where observations were taken every few days and errors were necessarily serially correlated it was shown how the assumption of uncorrelated errors was important for the comparison of different treatment means but unimportant for comparisons between different day means.

For testing equality of variances by the F -test or Bartlett's test the assumption of normality was very important, and Dr Box described how he had developed a test which approximated to Fisher's randomisation tests and which was independent of the type of distribution. This more 'robust' test was of necessity slightly less sensitive than the present tests, but the loss in sensitivity was remarkably small.

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

Reports of Meetings

BRISTOL GROUP

16th December 1954. Amount of Experimentation: D. H. Rees

This talk was substantially the same as that given to the Birmingham Group of the Industrial Applications Section on 6th January 1954 and reported in *Applied Statistics*, 3, 130.

20th January 1955. The Measurement of the Audience for Commercial Television: G. H. Dowson and R. Brook

The requirements of the consumers of such information were detailed, various methods of obtaining information were discussed, and their advantages and disadvantages were examined.

17th February 1955. The Use of Statistics in the Investment of Funds
G. H. Ross Goobey

Mr Ross Goobey mentioned the sources available for producing relevant statistics, their application to the framing of policy in investments, and the use of graphical methods of following fluctuations in the market. His remarks were illustrated by charts concerning long-term and short-term variations in particular stock items, together with explanations of the use made of such information.

BOOK REVIEWS

C. Fryer: *Elements of Statistics*. New York: John Wiley and Sons; London: Chapman and Hall, 1954. Pp. vi + 262. 38s.

It is easy to believe the author's claim that the material used in this book has been tried out in many classes and by several instructors for almost a decade, for although its scope is limited to an exposition for non-mathematical students of the simplest types of statistical technique, it is presented with a painstaking and sympathetic thoroughness which would make it an admirable guide to the lone worker. The student's queries are anticipated and answered, and he is given a rich supply of examples, designed to provoke thought as well as to exercise technique.

Topics covered are frequency distributions, probability, the binomial and normal distributions, and Student's t , with brief mention of quality control charts, the χ^2 test, and Spearman's rank correlation coefficient; an adventure outside the usual range of techniques is an exposition of an equivalent to the test using sample ranges.

A notable feature of the book is the number of results of sampling experiments given; the author claims, and I think rightly, that 'most students need this sort of empirical evidence before they really understand the nature and use of sampling distributions'.

There are one or two minor faults; the author allows his interest in blood groups for illustration of probability to run away with him to an unhappy degree of complexity, and is apt to be incomprehensible in some of his exercises on inheritance; but considered on the whole, though dear at the price, the book is an excellent introduction to the subject.

G. H. JOWETT

Ervin F. Schmid: *Handbook of Graphic Presentation*. New York: Ronald Press Company, 1954. Pp. vii + 316. \$6.00.

Professor Schmid has made a book of a subject that occupies only a few pages, if so much, in most textbooks on statistics. But the subject deserves, and needs, more attention from people who present statistics, and this book is worthy of attention. It deals with all forms of graphical presentation, including statistical maps and what are called 'projection techniques' for producing perspective diagrams that look solid. The 'know-how' of designing and producing diagrams is dealt with, and there are over two hundred illustrative figures. Altogether the book forms an excellent handbook for the instruction of the beginner and for suggesting ideas for effective presentation to more experienced statistical draughtsmen.

While perusing the book, however, I reflected on the paucity of knowledge to the extent to which graphical methods help in giving an understanding of statistical data. Can members of the public comprehend any but the simplest graphs? How far do the more pictorial methods of presentation—and they can be very elaborate—really stimulate interest in the substance, and how far do they provide distracting irrelevancies? Statistical draughtsmen should do some consumer research.

L. H. C. TIPPETT

R. L. Ackoff: *The Design of Social Research*. London: Cambridge University Press, 1959. Pp. xi + 420. 56s. 6d.

It is always something of a pleasure to encounter a textbook which takes a novel approach to its subject, even if the reading requires more effort than usual. Professor Ackoff's book is in many ways unconventional in approach; it is also in parts difficult reading.

His aim has been 'to provide a (not *the*) methodology of social research. That is, I have tried to raise to consciousness each type of decision that can arise in the course of research and to suggest methods by which these decisions can be made and criteria by which these decisions can be evaluated'. Professor Ackoff means what he says. Where other books are content simply to stress the importance of defining carefully any concepts used in the research, this one uses some ten pages to explain how to go about this task: what criteria to use for judging the soundness of a definition, and so on. Again, most writers in this field emphasise the need for establishing an order of priorities for the various objectives of a research; Professor Ackoff pursues this to its logical conclusion by proposing ways and means of ranking 'a set of objectives in a non-arbitrary way'. At times one may feel that he is making life unnecessarily complex, but his approach does have the great merit that it forces the researcher to think about choices he might otherwise have made almost unconsciously.

After an introduction, the book has two chapters on the formulation of the research problem and what the author calls 'the idealised research model'—the procedure the researcher might follow were he completely unrestricted by practical considerations. It is here that the selection of concepts, problems of definition, problems of quantification, and so on are dealt with. After this comes the core of the book, a series of chapters devoted to the statistical phase of a research design. The pattern is logical, beginning with a useful discussion of sample design, proceeding to the logic underlying statistical methods of drawing inferences from the sample, and ending with three chapters on tests of hypotheses and estimation. Throughout the author is less concerned with formulae and computational techniques than with underlying ideas. The chapters on tests of hypotheses are particularly valuable in their step-by-step discussion of some forty different tests. However, though these statistical chapters are written clearly and with considerable skill of presentation, it is difficult to agree with the publishers that 'a knowledge of statistics is not necessary for profitable use of this book'. Of course, it all depends on what one means by 'profitable'; but it is very doubtful whether the reader with no knowledge of statistics would be wise to make straight for this book.

The final chapters are devoted to the observational and operational phases of research design, dealing with questionnaire and attitude-test design, interviewing, response errors, pilot surveys, administration, cost and time schedules, personnel, and so on. Inevitably, these matters are presented more on a basis of opinions, common sense, and empirical experience than in rigorous scientific terms. The author rightly tries to avoid filling these chapters with apparently common-sense principles like 'observation should be made as carefully as possible'. In one sense such a rule is tautological but, in a more important one, is positively bad advice: everything depends, after all, on the cost of 'carefulness' and on the accuracy required in the research. The author (who in the last chapter is joined by L. Pritzker

manages to translate such statements into as helpful and precise rules as is possible. Altogether his book should prove widely useful. C. A. MOSER

Carl A. Bennett and Norman L. Franklin: *Statistical Analysis in Chemistry and the Chemical Industry*. New York: John Wiley and Sons; London: Chapman and Hall, 1954. Pp. xvi + 724. 58s.

While this book has only two authors it has a multiplicity of godparents. The Committee on Applied Mathematical Statistics of the National Research Council have sponsored the book to a remarkable extent, having observed some years ago in the chemical industry an increasing need for a book on applied mathematical statistics with examples from that field. Dr Bennett is a statistician for the GEC in the USA and Dr Franklin is lecturer in chemical engineering in the University of Leeds, a happy combination made possible by the financial support given by the Office of Naval Research and the cooperation of the Universities of Princeton and Leeds, which enabled the authors to spend two summer terms together at Princeton. Several familiar names from the American statistical field are mentioned as having given advice or guidance.

The result is of course a lengthy book; it is doubtful whether anyone asked for guidance would suggest omissions rather than additions, but the presentation is nevertheless a unified and consistent one. It is significant of the times that the book is avowedly mathematical; the authors claim that the mathematical standard has 'been selected to exclude material which would be beyond the range of the engineering or chemistry graduate' and that they write 'at a level between that of the well-known works on mathematical statistics and other texts which deal almost entirely with the application of statistical methods'. They have done this, not by any extensive omissions (in spite of a statement to the contrary) but by expanding the symbolic statements where advantageous, by the insertion of additional steps, and by including explanations of the significance of the symbolic results.

The chapters are well arranged from the student's point of view, especially for the newcomer to the subject. He is given in three chapters (64 pages) a detailed review, with some formulae, of the whole subject. This involves repetition but seems worth while in a book of this length. This introduction is worthy of a better title than Descriptive Statistics. Next come some 65 pages entitled Mathematical Machinery which deal briefly with distribution theory, a detailed consideration of the normal distribution, and the distributions of some sample statistics. This is followed by two chapters of approximately equal length on statistical inference and relationships between variables. Then come two lengthy chapters (150 and 120 pages), the first of which might with advantage have been divided, on the analysis of variance and the design of experiments. The three remaining chapters are short and deal with analyses of counted data, control charts, and some tests of randomness. Several subjects not usually dealt with extensively, like discriminant functions and transformations useful in the fitting of regressions, are given full treatment. The examples are almost exclusively chemical.

It must always be a matter of opinion just how much mathematics should be included in a book definitely intended for students of applications. To omit it is to give an impression of mystery while to include it can lead to masses of symbols which deter all but the most painstaking student. In this

book the student has not been spared even semi-invariants; if one may fairly assess the chemists' partiality for mathematics from the sample of statements made in the reviewer's presence, this book will appear formidable to most. Nevertheless there has been considerable interest in the methods among chemists, and they and the chemical engineers have here a book of reference which should see them through every application.

There is an unfortunate opportunity for misunderstanding in the paragraph on confidence intervals. The reader is told (the italics are the reviewer's) that 'we determine (*from the sample*) two values A and B such that there is a probability $(1 - a)$ that $A < \theta < B$ '; and, later, 'It must be emphasised that we do not make the statement that θ has a probability $(1 - a)$ of falling between the given limits'. Although the excellent device of the confidence diagram (Fig. 5.1) is employed in further explanation, the mistake is then made in connection with that diagram of saying that '*for given $\bar{x}$* the probability is $(1 - a)$ that the unknown mean of the population from which the sample was drawn is included between the limits'. A frequency definition of probability might have saved the situation by demanding a statement of what total of events the $(1 - a)$ is a fraction.

The book is well produced, the notation conventional and consistent throughout (except that 'Ave' is used for 'expected value'), and the mathematics well laid out. The paragraph heads are given in the contents list thus usefully supplementing the index. There is a list of 128 numbered books and papers referred to in the text. No expense being spared, the usual tables of percentage points for the significance tests are given to four decimal places and the normal distribution functions to six.

E. D. VAN RES

Norbert Lloyd Enrick: *Quality Control through Statistical Methods*. New York: Rayon Publishing Corporation, 1954. Pp. 72 (paper). \$5.00.

This glossy publication is designed for textile mills. Simple statistical procedures are introduced in a 'told to the mill man' style and their application is illustrated. Since the author has been a consulting engineer and presents 'live' examples, we may presume that the publication portrays the most advanced quality control practice in textile mills in the USA; this is perhaps its main interest. According to the 'blurb', the format of the work 'takes it out of the high price brackets'—few individuals on this side of the Atlantic, at least, will want to pay so much for so few pages.

L. H. C. TIPPET

Lewis A. Maverick: *Productivity, a Critique of Current Usage*. Published by the author, 70 So. Oakland Avenue, Carbondale, Illinois, USA, 1955. Pp. 30, paper. 50 cents.

The subject of productivity continues to call forth articles and pamphlets, some analysing the concepts and defining the various measures used, others of a more 'debunking' kind. The pamphlet under review is of the latter sort. To Professor Maverick 'true' productivity is the marginal increase in output of, say, a firm or an industry divided by the associated marginal increase in some resource such as labour, and he criticises many uses of global ratios such as total output divided by total manhours employed. He criticises notably the use of these measures in determining wages, a development that seems to have gone much further in the USA than in the UK. The criticisms are really against a naïve interpretation of such ratios.

Professor Maverick examines various attempts that have been made to sophisticate the global ratios of output to input by, for example, subtracting the value of raw materials from the value of the output or adding to the input the labour 'embodied' in whatever capital equipment is used, but he does not find salvation in this direction. He advocates, I think wisely, that in order to avoid misconceptions it is better to give any ratios full descriptive names rather than the short title of productivity. His concluding paragraph

'Possibly some economist who considers valuable the productivity studies that are currently being conducted by agencies of the American Government may explain their value to the public so convincingly as to persuade adverse critics. But until such support is offered, in clearer fashion than anything that has come forth so far, the worth of the studies remains in the dark shadow of doubt'.

This caused me to reflect on the uses of comparative 'productivity' studies that are currently conducted in industries in several countries. They may or may not be useful to economists, but they do stimulate, and provide some guidance for, executive action. Productivity measures—perhaps we should say output-input ratios—like many statistical ratios, can provide a kind of afford for investigation and action even when they do not measure well-defined concepts. Action may require separate justification, but it can be stimulated by the statistical study.

L. H. C. TIPPETT

J. Struik: *A Concise History of Mathematics*. London: Bell, 1954. Pp. xix + 299. 14s.

To compress the history of mathematics into three hundred pages, even when the last fifty years is omitted, is to attempt the impossible, as the author admits; but this said, the reviewer found himself quite fascinated at times by the material presented in this book, especially by the illustrations. These include portraits of many of the great mathematicians, and also reproductions of pages from some of the mathematical classics (for example, Euclid's *Elements*, and Leibniz on the calculus). If one adds that the history extends back to the number language of primitive races and traces the influence of Greek and Oriental mathematics on the later developments in Western Europe, some indication may be conveyed of the impression left with readers of this book, of mathematics as one of the supreme cultures and techniques, intimately related at each stage of its progress both to particular individuals and to the societies in which they lived.

Obviously the statistician interested in the use of mathematics in his own field should not expect to find very much, for statistical mathematics is in any case a mere three hundred years old. References—to the Fermat-Pascal correspondence on probability, to Huygens's book *De ratiociniis in ludo aleae*, to de Moivre's discovery of the normal distribution, to Laplace's monumental work *Théorie analytique des probabilités*, to Bayes's theory of inverse probability, to Gauss's method of least squares—are all there, but brief. (The reader should not be mystified by a reference on p. 192 to the use of the principle of initial velocities in statistics; for the latter, read 'statics'.)

As partial compensation for the brevity of much of the text, a list of some of the most important books on the history of mathematics is given in the introduction, as well as further reference lists at the end of each chapter.

M. S. BARTLETT

PUBLICATIONS RECEIVED

Richard Padley and A. B. Neale (Editors): *Abstract of Statistics, Number Three 1952-1954*. Birmingham Corporation, 1955. Pp. xiii + 139. 10s.

This excellently produced volume, modelled on the national Annual Abstract of Statistics, gives statistics for the City of Birmingham under the headings: population; vital statistics; health; care of children; miscellaneous social services; transport, communication, and water supply; employment; industry, and trade; government; meteorology. Only a few of the figures go up to 1954.

W. F. Harris: *Statistics and the Accountant*. London: The Incorporated Accountants' Research Committee, 1954. Pp. 38. 10s.

CURRENT TRENDS IN NATIONAL INCOME STATISTICS

CHRISTOPHER T. SAUNDERS

In this survey of the wide field of national income statistics Mr Saunders describes the various types of data collected and the uses to which they can be put. He also indicates their shortcomings and the opportunities for further work on this subject.

In most countries some attempt is nowadays made, officially or less often privately, to compile statistics of national income, output, and expenditure quarterly, annually, or from time to time. Such statistics may be described as estimates of national income and expenditure (as in the official *White Papers* and *Blue Books*¹ published annually in the United Kingdom, the reference giving the most recent examples), or as national product and expenditure; as social accounts; or as national income accounts. The various descriptions have traditional associations and undertones of special meaning, but the body of statistics presented—conveniently described here generically as ‘national income statistics’—is essentially similar. The object is to present as comprehensive a statement as possible of the structure and operation of a national economy so far as it can be expressed in terms of money, using either current prices or the prices of a base year.

Much interest attaches to the measures of aggregate national income, product, or expenditure, and to their changes over time. But the significance of changes in these aggregates (even when they have been adjusted to eliminate the effect of price changes) is rightly limited by philosophical doubt about the extent to which the ups and downs of millions of individual efforts and experiences can be combined in an algebraic sum. The gross domestic product—the total output of goods and services—of the fifty million residents of the United Kingdom, operating with a changing amount of natural resources and capital equipment, is stated to have increased between 1948 and 1953 by 15 per cent, valued at 1948 factor cost. Can one find philosophical justification for comparing this 15 per cent with the 20 per cent increase in country A (even assuming precise comparability of statistical concepts and perfection of statistical data) or with the 10 per cent increase in the United Kingdom in some previous five-year period? The economic philosopher may well answer ‘No.’ The practical economist may well answer ‘Yes: provided that its application is restricted to certain limited purposes connected with the laws and nature of economic growth, or with the detailed analysis of business fluctuations.’ Again, total consumers’ expenditure on goods and services in the

United Kingdom is stated to have increased by 5 per cent between 1948 and 1953, valued at 1948 market prices. Is this an appropriate measure of the change in 'welfare' derived from goods and services by the fifty million consumers (ignoring the changes in welfare derived from changes in leisure, weather, security, etc., which are not comprised in the statistics)? What increase in welfare was derived from the simultaneous increase of government expenditure and gross investment? Should not account be taken of any changes in distribution of income and, if so, how? If such a measure of changes in aggregate welfare were the main object of compiling the statistics, then the philosophical taxpayer might reasonably wonder whether the marginal utility of his tax payment is justified by the marginal productivity of the official national income statisticians.

However, the justification for the body of statistics of national income, product, and expenditure does not lie in the compilation of these massive aggregates for their own sake. The primary interest lies in the various elements that go to make up the aggregates and in the changes in their composition from year to year—in such matters as the contributions of different industries to national output, in consumers' expenditure on various items of consumption, in the relationship between investment and consumption, in the relation between income from work and income from property, in the contribution of different sectors of the economy to total saving, in the relation between the savings and investment of each sector.

From this point of view it may be thought that practically the whole corpus of economic statistics is relevant to the national income statistics even a variety of other statistical material, including births and deaths, temperature and rainfall, may help to explain the constituents of national income and expenditure. There is indeed no clear line to be drawn. There is no particular reason in principle why the *Blue Book* should provide index numbers of consumers' real expenditure on coal (derived from the number of tons bought) but not state the number of tons directly. The considerations determining what should go in, and in what detail, and what be excluded, are considerations of convenience as well as of relevance to the main themes.

National Income Statistics and Social Accounts

The trend in recent years has been to regard national income statistics as an organised system of national or social *accounts* rather than as a method of compiling and presenting estimates of aggregate income, output, or expenditure. The application of accounting concepts has been well described in J. R. Hicks's *Social Framework*² and in Richard Stone's appendix to the report of the Committee of Statistical Experts of the League of Nations³ (a post-war report of a committee set up just before the war). It is further developed in a new book, *National Income and Social Accounting*, by Harold C. Edey and Alan T. Peacock,⁴ which serves not only as an introduction to national income statistics but as

summary of what has been learnt in recent years from the use of these statistics, and of the new horizons now coming into view.

The thought behind the presentation of social accounts is simple enough. 'Economic activity' consists of a vast number of transactions among individuals, businesses, and governmental agencies within the country and outside it. Most of these transactions are recorded in the accounts which are actually compiled, in one form or another, for business control or for other reasons. Other transactions could be, and sometimes are, recorded (e.g. the personal accounts of the readers of this journal); they can in any case be reconstructed in the aggregate, with more or less approximation, from statistical sources. Still other transactions' are incapable of record. Some of the latter group, such as the annual value derived from owning a house, are normally included for the sake of comprehensiveness. It is desirable to avoid such statistical absurdities as a fall in national income because more houses are owner-occupied and fewer rented; some approximate measure is then substituted, for example the rents of comparable rented houses or Schedule A values. Other 'unaccountable' transactions are best ignored because of the arbitrariness of valuation that would be involved; the classical case is the value of the services of housewives.

The essence of social accounting is to provide a manageable system of classification of the transactors and of their transactions which brings to light those changes in economic activity that are important for economic analysis, and ultimately for economic and social policy and political judgement.

Sectors of the Economy

The fifty million 'transactors' or economic units in the United Kingdom must then be classified into a number of groups, described as the 'sectors' of the economy. The object is to find groups each of which is, very broadly, homogeneous in economic behaviour and distinguishable in economic activity from the rest. What should be a subsidiary consideration is sometimes predominant: namely that the groups should conform with the categories which are distinguished in the statistical and accounting information that is available. In the British statistics six sectors are distinguished: persons (meaning, broadly, people considered as the ultimate recipients of income and as the ultimate consumers); companies (the major part of private business); the public corporations (the nationalised industries); the Central Government; local authorities; and 'the rest of the world' (the latter being treated as a 'sector' for recording transactions between UK residents and foreigners).

(Other meaningful divisions might be, and sometimes are, used. For some purposes, banks, insurance companies, and other companies whose main business is transactions in money and claims to money, must be distinguished from companies whose business is to sell goods and services; such a distinction is essential, and is commonly made, when

we are concerned with the relation between credit flows and business activity. Other distinctions could usefully be made: for example, the personal sector could profitably be divided between rich and poor, or between young and old, if we are concerned with propensities to save and consume. Statistically, this particular distinction can be drawn only by actual collection of personal accounts through sample budget surveys. Again, useful conclusions might be drawn about the behaviour and motivation of business enterprises if the transactions of large companies could be distinguished from those of small companies.

Still other distinctions would have theoretical value if they had practical sense and if accounts or statistics could be obtained. A troublesome example is the treatment of personal businesses. It is one of the awkward facts of life that a sizeable proportion of total national output is derived from unincorporated, often one-man, businesses—for example, a large part of farming and of retail trade. The proprietors may be able to produce trading accounts, and their income is assessed to tax under Schedule D as a form of profit; from sample surveys, estimates of the consumer expenditure of the proprietors could be derived. But how much of the unspent income is personal saving, analogous to the saving of an employed person? How much is business saving, analogous to the undistributed profits of companies? The proprietor himself could not necessarily answer the question, and if he attempted an answer it would be no more than an arbitrary and subjective allocation. Should the surplus of income from such businesses over current expenditure be treated wholly as personal savings, or as savings of a possible 'business sector' comprising both corporate and unincorporated enterprises, or divided between them? The British official statistics allocate the whole saving to the personal sector, but this is not the only possible solution. This is not a problem which might be solved by further statistical investigation (as Edey and Peacock seem to imply that it might be). It is a fact of our economic structure which exposes the inevitable artificiality, in quite important marginal cases, of any classification of economic units. It is also one of the facts which sometimes make it impossible to fit statistics to the categories of economic theory—in this case, the theory concerned with the behaviour of 'entrepreneurs'—which for several reasons cannot be accommodated to any imaginable body of statistical or accounting facts.⁵

Broadly, however, the system of classification by sectors serves its purpose of revealing the interaction between one branch of the economy and another. An increase of £ x in retirement pensions, for example, can be traced in two directions. It shows up as an increase of £ x in Central Government expenditure (Table 4 of the *Blue Book*); it reappears as an increase of £ x in the receipts of the personal sector (Table 2, income side); it is reflected in an increase, not precisely identical, in expenditure of the personal sector (Table 2, expenditure side); it is reflected, with many similar changes, in a rise in output and therefore, perhaps, in a rise in employment income and company

profits. In the other direction it results, other things being equal, in a fall of £ x in the surplus of Central Government on current account (Table 4) and thus in the contribution of Central Government to total national saving (Table 6).

Classification of Transactions

The classification of transactions is in part derived directly from the categories used in commercial accounting. There are two important distinctions to draw. The first is that between current and capital transactions. For economic analysis, the basic distinction here is that between the two fundamental objects of economic activity: providing for present needs and providing for future needs, broadly described as consumption and investment or adding to wealth.

The second distinction is less fundamental and familiar and is partly due to the special requirements of social accounting arithmetic. Three categories are here distinguished. (a) Transactions in goods and services—normally the exchange of money for a commodity or service. It is from the record of these transactions, whether current or capital, that the total output and expenditure of the nation, and the composition of these totals, must be compiled. (b) 'Transfers', which represent payments of money against no corresponding provision of goods and services. There are many kinds of transfers, some governed by a form of contract, some not. Examples are payment of retirement pensions and national insurance benefits; payments of interest on National Debt; payments of war damage compensation; grants by the Central Government to local authorities; grants by one government to another, such as Marshall Aid or Colonial Development grants. It is important to realise that such transfer payments may not be wholly unilateral, or unrequited'; pensions, for instance, might be regarded as deferred wages. The essential point is that the money payment cannot sensibly be treated as a measure of the amount of 'service' provided within the accounting period, or to be provided in the past or future. (c) The third kind of transaction is transactions in money claims—primarily the borrowing and lending of money. Such transactions are by their nature capital transactions. They resemble transfers in that there is no corresponding provision of goods or services. They differ from transfers in that transactions in money claims affect by a measurable amount the shape of future accounts. Thus loans by the Central Government to local authorities result in a measurable amount of interest and repayment, and a change in the outstanding amount of local authority liability and Central Government credits. A capital grant might have precisely the same effect on current economic activity (the building of such and such schools), but would not measurably affect the future financial relationships between the transactors.

Social accounting thus represents a diversion of interest from the building up of aggregates to the display of relationships between parts of the economy. Compilers of national income estimates, from

Gregory King in the seventeenth century onwards, have always been concerned with problems of distribution as well as with aggregates. But their aim has been to record only those transactions which contribute to total national income (or output, or expenditure). The development of social accounting has, so to speak, 'grossed-up' the statistical records presented in national income tables. Thus a table showing the composition of total national income need show (as does Table 1 of the *Blue Book*) only a single figure of total company trading profits before tax. But the table showing the current account of the company sector (Table 23) reveals how total profits were apportioned between interest and dividend payments, taxes, and the balance which is company saving. A part of the interest and dividend payments reappears as a constituent of personal income (Table 2), while the tax payments reappear in the income of Central Government (Table 4), and the saving reappears in a combined capital account for all sectors (Table 6). Again, transfer payments to persons (such as retirement pensions) would not appear at all in a table showing the composition of national income, since they do not represent payment for any current contribution to output. In the social accounts they are shown as Central (or local) Government expenditure and as part of personal income.

If all transactions can be fitted into this framework of sector accounts, the result is a picture in accounting terms of the variety of economic activity. It is of course true that the accounting framework—like any description in the language of numbers—is a strait-jacket. Much that is important disappears from view. Many unimportant features acquire unnecessary prominence because the framework happens to provide cells for them.

Social Accounts and National Income Statistics

The art, as distinct from the mechanics, of presentation is to produce a set of tables, built with the aid of this social accounting scaffolding, but from which the scaffolding poles have been removed. Some understanding of the rationale of the accounting system is indeed necessary for refined analysis of the information provided, just as some understanding of commercial accounting is needed to grasp the full significance of a company's accounts. But the tables should be designed primarily to convey essential information about the economy. Many of them are perfectly comprehensible to a reader who has no understanding of, nor interest in, the social accounting system. Statistics of the total wage bill; of personal income and current expenditure and saving (granted the difficulties of definition, which are not due to social accounting as such); of transfer incomes received by persons from public authorities; of net government borrowing or lending; of fixed investment in buildings or machinery—indeed, most of the items appearing in the *Blue Book*—are intelligible in themselves, and in their general relationship to each other, to anyone with some grasp of what economic statistics are about. The reader must also, of course, be aware of a

handful of notorious 'catches': he must know, for instance, that all forms of income and fixed investment are shown before deduction of provision for depreciation, the broad significance of stock appreciation, and that taxes on dividends are regarded as falling on the recipient and not on the company that in fact deducts tax before payment.

If the object is to convey essential information about economic activity, the provision of fully 'articulated' social accounts, in which each inter-sector transaction is classified both according to the class of transaction and according to the two transacting sectors, may be a hindrance rather than a help, even to the experts in social accounting. For these reasons the tables in the British *Blue Book*, although built with social accounting scaffolding poles (and perhaps still showing traces of them here and there), are primarily intended to be tables showing what the reader wants to know about economic events.

Input-Output Tables and National Income Statistics

The development of input-output tables, or accounts of inter-industry relationships, associated with the name of Leontief, can usefully be regarded as another stage in the 'grossing-up' of the national accounts. It is so treated in Edey and Peacock's book,⁴ where the construction of a simple matrix of input-output transactions is very appropriately worked into the elementary statement of the building up of national accounts. An input-output table is essentially another way of summarising the complex of transactions between the various parts of the economy. But while the social accounts are designed to reveal transactions between more or less homogeneous groups of economic units (persons, companies, public authorities, and the other sectors), the main interest of the input-output table lies in transactions between *industries*. The input-output table shows how each industry purchases materials from a variety of other industries, adds its 'net output' measured by wages, salaries, and profits, and resells the composite goods or services to another industry or to the final purchaser.

The main object of an input-output table is to make it possible to trace a change in final expenditure on goods and services to all the industries directly and indirectly affected, including consequential changes in imports. The difficulty about the use of input-output tables is that the relationships vary with technical change, changes in relative prices, and changes in consumer preferences. Hence one cannot be wholly content with a description of the inter-relationships that held some years ago. Yet the labour of constructing a complete matrix of relationships, if it is at all detailed, is so heavy, and the amount of information required—much of it not regularly available—so great, that it is always difficult to be confident that the conclusions drawn apply to the current situation.

Prediction and 'Model' Building

In one way or another, the main uses of national income statistics,

perhaps of most economic statistics, are bound up, explicitly or implicitly, with prediction. Statistics for the past may have a value of their own, to assist understanding of the past from the standpoint of the historian, or to justify or criticise those responsible for past events. But the major practical use is to draw conclusions about events, and appropriate behaviour, in the future—and, in certain circumstances, to control future events.

Such prediction may take the form of forecasts expressed in statistics relating either to individual series or to 'models' exhibiting a whole system of internally consistent relationships. Others may prefer to concentrate on a description of the forces operating on, or in, the economy in the more or less immediate past; the reader is left to draw the conclusion, if he so wishes, that the same forces will operate, with greater or less strength, and will have the same effects, in the more or less distant future.

Few forecasts—whether quantitative or qualitative—are explicit forecasts in quite the same sense as those of the Meteorological Office. The meteorologist is obliged, for public presentation, to express what is only a probability in a fairly definite statement. He can hardly say to the public: 'If this anticyclone off the Azores continues to move north-east, without breaking up, which is about 75 per cent probable, it will be fine and warm in south-west England on Tuesday.' The economist is slightly more fortunate. He can guard himself by making his assumptions explicit. '*If* events in the outside world result in a rise in commodity prices of X per cent, *if* wage and salary rates in the United Kingdom rise by Y per cent, *if* last year's profit : wage ratio remains constant, and *if* tax rates are unchanged and no change of policy occurs affecting public authorities' expenditure, *then* imports under half a dozen categories will be so much, consumers' expenditure, exports, and investment under the main headings will be so much each, the index of industrial production by broad industry groups will be as follows, and the balance of payments such and such.' It is true that the assumptions chosen since these are chosen and not others, are likely to be taken to represent the author's 'best guess' at what is likely to happen. But even apart from this, such forecasts have their value. It is not obvious at first sight what are the consequences throughout the economy of primary assumptions about 'exogenous' forces such as those suggested in the illustration above. A rise in import prices or wage rates, for example, works its way deviously through the system, with time lags which it is worth while trying to ascertain from previous experience. Again it is impossible to determine by intuition or general knowledge, even in orders of magnitude, what will be the effect on the output of individual industries of given increases in consumers' expenditure, investment, exports, or government expenditure. Even the simplest form of input-output table can play an important part in bringing to light, in broad terms, how different industries may be affected by given changes in final expenditure. The need for tracing such secondary consequences, rather than

the hope of forecasting the level of wage rates or the course of commodity prices, is one modest aim of practical model-building.

It will be evident to the reader of Shankleman's article on 'Econometrics'⁶ and Leser's article on 'Elasticities of Demand'⁷ that much of the analysis required to determine the laws of behaviour of the economy is still highly experimental and tentative. The analysis of consumers' demand, in particular, suffers badly from lack of data from which generalisations can confidently be applied to the present time. A long period of world shortages, rationing and price control, accompanied by substantial changes in the distribution of disposable incomes, might obviously play havoc with conclusions about consumer demand derived from experience of a previous period. But nobody can be certain that the determinants of average behaviour have been affected, and the test of some of the work already done on demand analysis may come quite soon.

Financial Assets and Liabilities

One of the major incentives to the development of national income statistics in their present form was the crucial importance attached by Lord Keynes to the factors determining the propensity to save and the propensity to invest. These are fields in which the amount of statistical analysis so far done is in fact extremely small. The difficulty again is lack of material, and especially of material relevant to present economic conditions.

Two promising lines of development of the study of saving are now being opened up. The first is the study of personal savings through the sample surveys of households conducted by the Oxford Institute of Statistics over the past three years. The first of these surveys, relating to savings in 1951-52, has been described by Lydall.⁸ The surveys bring to light information about changes in various kinds of assets and liabilities, analysed by income groups, occupational groups, composition of household, etc. The second is the analysis of the financial accounts of public companies instituted by the National Institute of Social and Economic Research.⁹ The intention is to analyse both the profit and loss accounts and balance sheets of companies in the form of statements 'Sources and Uses of Funds'. Broadly, this form of analysis will bring to light, mainly from the published accounts of companies rearranged in a standard presentation, how far companies in industry as a whole, and in individual trades, draw on outside sources of finance and how far on internal funds, and the relationships between investment in fixed assets, stocks, and financial assets on the one hand, and their sources of finance on the other. The studies should make it possible to draw some conclusions from current practice about the propensity of businesses to save and to invest in the various kinds of assets.

These two studies illustrate a comparatively new development of national accounting. Great emphasis has been put, for some time, on the study of income and expenditure from the point of view of the supply and demand for, real resources—goods and services for consumption

or for adding to the productive capacity of the nation. In times of shortage of real resources, and of potential or actual excess of demand for real resources over supply, this was indeed the use of national income statistics which had most importance both for policy and for analysis. It will continue to be a major use, and there is still plenty of opportunity for the improvement of the statistical material required for this purpose.

At the same time, greater interest is now taken in the largely complementary study of 'money flows', or 'asset structure' to which Edey and Peacock⁴ rightly attach great importance. The most useful application of this line of study is probably the analysis of the capital accounts of the various sectors of the economy—of changes not only in physical assets but in financial assets and liabilities (the various forms of borrowing and lending). The need is felt for more knowledge not only of changing levels of fixed investment and of investment in stocks, but of how, and from what sources, this investment is financed—of who is lending to whom.

The social accounts reveal, for each sector, an excess or deficiency of money income (after tax and other transfer payments) compared with the sector's total expenditure, current or capital, on goods and services. The *Blue Book* tables show this excess or deficiency, but only as a single net figure for each sector (described in Table 45 of the 1955 *Blue Book* as 'net acquisition of financial assets plus net overseas investment'). From the information now available, it is in fact impossible to provide more than this single figure, which is essentially a residual derived from the balance of known transactions. One major object of the study of money flows is to gross-up, in as much detail as possible, this net figure, so as to show for each sector the gross borrowing and gross lending, both long-term and short-term. The objective would be a matrix revealing a sector's borrowing from, and lending to, each other sector.

It would not be easy to develop *comprehensive* statistics from which accounts of inter-sector money flows on these lines could be built into a complete matrix, properly integrated with the national income statistics. But much progress can be made by building up from sample surveys of households, and from analysis of as many company accounts as possible, on the lines now being pursued and described above.

Changes in Capital Assets

The analysis of investment in fixed assets, and of the factors determining the propensity to invest, has, not only in this country, been impeded by the difficulties of dividing total investment between the replacement of existing assets and net additions to assets. This important distinction has not been made in the *Blue Books*. There are great difficulties here, familiar to accountants as well as to economists and statisticians, because it is impossible to determine in any accurate way how far the installation of a new asset represents simply the replacement

an old one, how far an addition to the productive capacity of the enterprise. Buildings and machinery last a long time, and when they come to be replaced the new installation is rarely identical in design or capacity with the old. To this permanent problem of the measuring 'replacement' is added the even more serious difficulty caused by the great change in prices over the life of most existing assets. So long as prices are stable, it is possible to make use of the conventions of accountancy, and to regard as net investment any excess of total expenditure on capital account over the provision for depreciation. The implication, which is not indefensible when prices are broadly stable (although it is still highly arbitrary), is that assets costing £ x today can be regarded as 'replacing' assets of the same description costing £ x twenty years ago. When prices change, the same convention can be adopted, but allowance made for price changes by means of price indices. This problem involves estimating the distribution of existing assets according to their period of installation, and applying price indices to the installations of each year so as to raise their original cost to current values. The difficulty of constructing price indices relating to capital goods and extending over long periods is notorious, just because of the changes in quality referred to above.

The elaborate calculations required to carry out this exercise have recently been done for the United Kingdom by P. Redfern.¹⁰ It should be understood, as Mr Redfern observes, that there is no perfect statistical solution to the problem of 'measuring net investment; there can only be approximations and arbitrary conventions'. The approach outlined above is essentially the accountant's approach, but with adjustments for changing prices. An alternative approach is the direct measure of changes in productive capacity. No existing statistics can shed much light on this (except in a very narrow range of industry), and it would probably involve some very extensive research, as well as a good deal of consideration of just what is being measured.

It has been pointed out that the propensity to invest is in part determined not by the current cost of replacing existing plant, but by the extent to which assets have in fact been written-down in the books of the enterprise. The amount of depreciation actually provided by conventional accounting helps to determine the conventional amount of net profit, and therefore the apparent financial position of the enterprise as shown in its balance sheet. Precisely recorded facts, even though not directly relevant to the problem in hand, have their influence on decisions. To this extent the amount of depreciation at original cost—however it may be criticised as a measure of any economic concept—has a certain importance of its own. Since the amount of depreciation at original cost also helps to determine tax liability, it is, again, a figure of some significance in its own right.

Conclusion

Some examples have been given of the current lines of development

of national income statistics. As a broad, possibly too broad, generalisation it may be said that the main framework of descriptive statistics of income, output, and expenditure, and the methods of measuring the allocation of real resources, have been fairly well established in the United Kingdom and in most other countries with well-developed statistical systems. There are still many weaknesses and gaps in the data and in their treatment and presentation; in particular there are many difficulties in making estimates which are reliable enough, and can be compiled quickly enough, to have maximum value for government or business policy. But these can be regarded primarily as matters of detail, and the opportunities for substantial new work seem to lie in two other directions. Firstly, there is the almost undeveloped area of the study of changes in the financial assets and liabilities of the various parts of the economy. The first problem to be solved here is the prosaic task of collecting the relevant information. Secondly, there is still ample scope for analytical study of the existing statistical material for analysis of the laws of motion of the economy. Too little is known, for example, about the changing inter-relationships between income and the composition of expenditure, and saving; about the incidence of demand changes on the industrial structure; or about the relation between changes in the capital stock and changes in output.

REFERENCES

- ¹ *National Income and Expenditure 1948 to 1954*, Cmd. 9423; and *National Income and Expenditure, 1955* (the national income *Blue Book*). HMSO, London.
- ² HICKS, J. R. (1952). *The Social Framework* (2nd Ed.). Oxford University Press.
- ³ UNITED NATIONS (1947). *Measurement of National Income and the Construction of Social Accounts*.
- ⁴ EDEY, H. C., and PEACOCK, A. T. (1954). *National Income and Social Accounting*. Hutchinson's University Library, London.
- ⁵ BURDETT, D. K. (1954). 'Social accounting in relation to economic theory', *Econ. J.*, **64**, 256.
- ⁶ SHANKLEMAN, E. (1954). 'What is Econometrics?', *Applied Statistics*, **3**, 85.
- ⁷ LESER, C. E. V. (1954). 'The measurement of elasticities of demand', *Applied Statistics*, **3**, 74.
- ⁸ LYDALL, H. (1953). 'National survey of personal incomes and savings, Part IV', *Bulletin of the Oxford Institute of Statistics*, **15**, 341.
- ⁹ LUBOFF, A. (1953). 'The finance of public companies', *Accounting Research*, **4**, 285 and 291.
- ¹⁰ REDFERN, P. (1955). 'Net investment in fixed assets in the United Kingdom, 1938-53'. *J. R. Statist. Soc.*, **A**, **118**, 141.

FUNDAMENTAL PROBLEMS IN THE SUBJECTIVE APPRAISAL OF FOODSTUFFS*

ROLAND HARPER

Department of Psychology, University of Leeds

Although subjective methods are widely used in appraising foodstuffs, much work remains to be done in studying the problems involved. In this article Dr Harper gives a general survey of the field and indicates some of the difficulties. He also examines some possible reasons why in some quarters subjective methods are thought to be always inferior to objective measurements, and makes a strong plea for a re-assessment of the position.

Introduction

The many problems posed by the subjective appraisal of foodstuffs inevitably involve many different branches of human knowledge. Food science requires the investigator to blend information drawn from physics, chemistry, statistics, and psychology, to mention only a few of the relevant disciplines. Each has its own special terminology, and in order to be generally understood a great deal of 'translation' into everyday language is essential. However, this inevitably increases the length which is taken up by the same amount of information compared with its statement in purely technical terms. Obviously, a comprehensive treatment of the many problems raised by the subjective appraisal of foodstuffs might well occupy a whole volume rather than a single article. In his monograph on *Flavor*, Crocker³ has surveyed the problems of food grading in general terms. Comprehensive reviews on particular aspects of the appraisal of foodstuffs have appeared in *Advances in Food Research*, a periodical now in its fifth volume. The writer himself has prepared three separate reviews of relevant literature during the past six years. He has also recently surveyed from various angles the question of food preferences and their complex origin (see Harper^{9, 10, 12, 13}). Attention should also be drawn to a bulletin prepared by Dawson and Harris⁴ on behalf of the US Department of Agriculture. Although the quantity of American work now appearing in this field is considerably greater than that which is being carried out in this country, there is no reason to believe that we are lagging behind in quality of work. Perhaps the present survey may help to stimulate greater activity in this field in the United Kingdom. Changes in conditions since the ending of food rationing have turned what used to be

* This article is an expanded version of part of a paper given to the Industrial Applications Section of the Royal Statistical Society (London Group) in December 1954.

studies of primarily academic interest into a rich source of guidance on urgent practical problems. However, effective application of what is known can be achieved only by properly planned investigations within the food industry itself. From time to time it may be profitable to take practical problems back into the academic laboratory to examine them further, but such investigations are no substitute for work sponsored by the industry itself and geared to its own recognised problems.

The remainder of this article falls under the following headings:

Functions of the subjective appraisal of foodstuffs.

Methods employed for the subjective appraisal of foodstuffs.

Types of judgement.

The effects of training and experience.

Other factors influencing human judgement.

Why have subjective judgements been under-rated?

Instrumental tests versus human judgement.

The examination of multi-dimensional relationships.

In view of the width of the range of topics considered no attempt is made to give a formal 'summing up'. Instead, the whole treatment has been carefully subdivided, which should make it easy for the reader to refer back to any particular item of interest.

Functions

The most effective way of considering the functions of the subjective appraisal of foodstuffs is simply to list a representative selection of them. Subjective appraisal is carried out with the following intentions:

(a) To provide essential information about *changes* in foodstuffs which can be detected by man. These changes may arise naturally, or they may be due to alterations or differences in manufacturing processes.

(b) To enable the investigator to identify the various characteristics which distinguish one sample of a particular product or food from another.

(c) To indicate whether 'quality' may be represented in terms of a single index or whether it has several independent 'dimensions'.

(d) To help to establish, within specified limits, a uniform standard of quality for both natural and manufactured products.

(e) To assign the various products to their relevant quality grades. These will largely determine the price and destination of a particular batch or quantity.

(f) To provide information which shows the relation between various 'objective' tests on a food product and the consuming individual's experience. Such information may be essential for the proper interpretation and use of objective test data.

(g) To provide information about the discrimination, acceptability and preference of various groups of consumers for different samples of the same product.

Ideally, it would be advisable to elaborate further each of these sub-headings. However, relevant comments will be made incidentally in other sections. Let us therefore turn to consider the main methods of subjective appraisal.

Methods

The methods in common use for appraising foodstuffs fall into three main categories. These are:

The laboratory panel.

Commercial grading techniques.

Consumer research techniques.

The Laboratory Panel. The laboratory panel may consist of persons either with or without extensive technical knowledge of the product concerned. The availability of personnel is a prior factor in the selection of a laboratory panel, although it is obviously desirable that the persons included shall at least be capable of giving reliable or consistent results. How such persons should be selected and trained does not concern us directly here, although such questions must be included among the fundamental problems facing the investigator. The detection of differences arising from changes in raw materials or processes is one of the special functions of the laboratory panel. Where the laboratory panel is not primarily concerned with how a particular product will meet the requirements of the consumer, there are occasions when this task may be taken over by the laboratory panel. Judgement will then involve some sort of mental standard of what the consumer requires. Such a judgement may be additional to the direct comparison of one sample with another which is normally made in order to maintain comparatively uniform quality in the face of changing conditions. In some branches of the food industry commercial grading is the task of a special group whose work will be considered next.

Commercial Grading Techniques. Grading is perhaps most common in those branches of the food industry which involve the least amount of 'processing'. Such a statement involves some reservations, for grading of cheeses is well established, and yet their manufacture certainly involves a considerable number of 'processes'. The task of the grader is to assign substantial quantities of a particular product to one of several categories or grades which determine price levels within those current at the time. Obviously there are many important problems involved in trying to make sure that the sample appraised is typical of the whole bulk. There is some doubt whether the problems involved are purely statistical. Techniques employed vary considerably, depending on the product and the traditions of the particular branch of the food industry concerned. In certain cases well-established and

statistically satisfactory sampling techniques will be employed. In other cases the selection of the particular samples to be appraised will be left to the *choice* of the grader; the task may be made easier for him by placing batches of similar origin close together (e.g. in cheese grading). In many branches of the food industry there exists a tradition for appraising separately a number of distinguishable characteristics. There are various methods of arriving at some indication of 'overall quality'. In some cases a separate overall assessment is made. In others some relatively untested logic is used and the scores for separate characteristics are added together so as to give a single index of overall quality. Most of these procedures are custom-built rather than scientifically tested. In the absence of evidence that a complicated system is necessary there is undoubtedly much wisdom in avoiding unnecessary untested and unrealistic refinements.

Consumer Research Techniques. Hitherto, consumer research on foodstuffs has received most attention in connection with manufactured products. There are many important distinctions between consumer judgements, those of the laboratory panel, and those of the commercial grader. With the exception of a few epicures, most consumers are unlikely to show unusually high discrimination, nor will they be in possession of much relevant technical information. Harrison¹⁴ has gone so far as to recommend that the limits within which quality control should be maintained in the food industry (American) should be determined by precisely those differences which the consumer rather than the expert can just detect. It will be obvious that the direction in which *preferences* lie must be considered as well as the detection of *differences*. However, it is unlikely that a regular pattern of preferences could emerge unless the group is able to tell the difference (discriminate) between the various samples. If one had to take into account *all* the distinguishable sub-groups of preference and opinion, then the task of consumer research would become impossibly complicated. The question is one in which specific aims must be taken into account. Broad and relatively superficial surveys may provide answers to one set of questions. However, there are occasions when such surveys must be supplemented by the more detailed information which may be provided only by the intensive study of small and relatively homogeneous groups of persons. On such occasions there may be much to recommend the use of interviewers who are not only specially trained in interviewing techniques but who are also well informed and interested in the problems under investigation. For further information on the many issues raised by consumer research in the food industry the reader should consult a recent review by Morse¹⁹ entitled 'Rationale for Studies of Consumer Food Preferences'.

Types of Judgement

More than one authoritative writer has suggested that judgements on foodstuffs should be divided into two types. Whether or not this

adequately covers the whole range of possible types does not concern us at the moment. At least it is relevant to draw attention to the distinction between the *grading* type and the *analytical* type of judgement, which reflects the distinction between the work of the professional grader and the laboratory panel. As we have seen, the grading type of judgement is concerned with the classification of samples or batches of a particular product into broad categories. Usually this type of judgement is made without an immediate comparison standard, that is, one that is actually present. However, some form of mental standard is implied, based upon technical knowledge, intimate familiarity with the particular product, and general past experience. In contrast, the analytical type of judgement is primarily concerned with the detection of changes or the detection of differences, and normally involves comparison of one sample with another, which serves as a standard. In certain circumstances the distinction between the analytical and the grading type of judgement is hardly maintained. Sometimes it is possible to provide direct standards to represent the main steps along the appropriate 'quality dimensions'. Obviously there is much to recommend the practice of providing such direct comparison standards. They would help in developing effective training methods and in the evaluation of existing grading practice. If such comparison standards were employed in grading and in the work of the laboratory panel then the functions of these two groups of persons could be practically indistinguishable. Each would have the task of matching a given sample to the corresponding one on a 'scale' covering the main sources of variation between the samples. (Many examples could be given of this approach to the standardisation of subjective judgements, although few would be taken from the food industry: e.g. the Munsell Colour System, the Crocker-Henderson Odour Classification Set.) Obviously, where several characteristics have to be taken into account several such scales would be necessary. Where complex qualities are involved judgement is likely to require some form of mental abstraction. Of course, in day-to-day commercial grading the regular use of such standards would be unlikely, and the various scales would have to be 'carried in one's mind'. However, where such standards have been developed they will have an important function to perform in cases of uncertainty or dispute. Furthermore, such standards are essential if deliberate and systematic training procedures are to be developed.

Training and Experience

Many important questions concerning the effects of training and experience could be discussed. The distinction between training and experience is a necessary one. Training refers to steps that may be taken deliberately to increase the effectiveness and the rate at which an individual assimilates new knowledge or new techniques. Experience refers to the cumulative effect of exposure to events or occurrences

without any systematic attempt on the part of another person to focus attention upon what is most important. Hitherto, deliberate training in the assessment of foodstuffs has been comparatively rare, although no doubt this state of affairs is changing. Indeed, systematic training is coming to be regarded as one of the most important needs of contemporary industry. In the present stage of development, clear evidence of the effects of training and experience on judgement will have to be drawn from outside the food industry. However, let us examine the types of problem within that industry which require attention.

As we have seen, subjective judgements seem to fall into two main categories, namely comparative judgements and those made without an immediately present comparison standard. The second group are sometimes referred to in psychological literature as 'absolute judgements', but this is really a misnomer because there is reason to believe that some form of mental standard is involved. An important question is: to what extent are such standards stable and unchanging? However, only rarely will current practice 'in the trade' be carried out along lines which immediately lend themselves to systematic investigation and the necessary statistical analysis. From one point of view, satisfactory investigations often require changes to be made in long-established practices. In view of this, if such studies show up the 'expert' in an unfavourable light, then he would be quite within his rights to demand that profound modifications in the tasks which he usually undertakes can be used to give valid information about his own special skills. The onus of satisfactory proof is on the investigator and not on the person tested. Even the most sceptical scientist must admit that the expert in the trade has his own special skills, yet these often seem to slip through the mesh of the net of systematic investigation. To give two examples there is every reason to believe that the pre-war cheese factor was able to direct the available products according to the known preferences of particular regions and sometimes according to the preferences of particular customer. Secondly, it was believed that there were definite regional preferences for different blends of tea and that these could be met by the trade. Yet when the trade expert can be persuaded to submit to carefully controlled tests the results often seem to suggest that he is not as good as he claims to be. However, we must recognize that the artificial and unusual circumstances surrounding such tests can largely invalidate them. If the information is to be taken as sound evidence, then either it should come from persons whose livelihood does not appear to be threatened by the aims of the investigation or it should come from everyday practice.

Can any relevant information about the effects of training or experience on subjective judgements be drawn from other fields of investigation? The writer carried out a number of investigations in which the persons tested were required to assess the difference between samples of india rubber by compressing them between finger and thumb. This task is not irrelevant to the assessment of foodstuffs for

texture and consistency, and other 'properties' are often assessed in this way. Only one example will be introduced here (see Harper,¹¹ p. 48). Measurements were made on each rubber sample, using a British standards Hardness Tester, so as to provide some information about the objective differences between samples compared. By the use of techniques which need not be described in detail it was possible to determine what physical difference between two samples was necessary for them to be assessed correctly (in terms of that physical criterion) seventy-five times out of a hundred. This is quite a common technique, which involves making use of the confusion of judgement as a measure of sensitivity. Under favourable circumstances *certain individuals* could be trained effectively. For example, one person showed a fourfold increase in sensitivity as a result of an hour's systematic training on alternate days over a period of about eight weeks. Thus at the end of the training period this individual was able to detect correctly, with the same level of success, differences which were only one-quarter the size of those needed in the untrained state. In addition, there were certain fluctuations superimposed on the progressive improvement which might well be related to mental preoccupation. It should also be noted that in the final state differences of about four-and-a-half per cent were being detected correctly seventy-five times out of a hundred. Since the term 'correct' is linked to the instrumental measurements it is also important to note that this level of sensitivity is roughly of the same order of magnitude as that of the accuracy of the physical measurements involved.

More recently McKennell¹⁸ has been able to exploit the fact that in the wool textile industry there is in existence a scale of Wool Quality Numbers which forms the basis of the buying and selling of wool tops. Wool tops are made up of combed wool fibres in the form of thick 'clips'. McKennell was able to examine both the ability of various people to match an individual sample against a scale formed by a number of other samples and the ability of these same persons to assign similar numbers without using a comparison standard. This is precisely the sort of task which, it has been suggested, should be developed in food investigations. Note that in this special case the investigation has been fashioned round the traditional features of the industry concerned. As a result of a series of tests, once a week, over a period of about eight weeks, a group of students showed a considerable improvement in their ability to carry out each of these tasks. In the case of the matching test the average error was reduced from two to slightly over half a step on this traditional Wool Quality Number scale. The judgements made without a matching standard were slightly less accurate, with an average error at the end of training of one and a quarter steps. Individual differences were marked, and one individual seemed to be just as good without as with the comparison standard. In order to provide some indication of the size of the changes referred to, it is worth noting that the most important single physical factor upon which

the Wool Quality Number depends is average fibre diameter. The steps on the traditional scale *normally* involve a difference of between five and ten per cent in average fibre diameter. Differences between different laboratories in measuring the average fibre diameter samples drawn from the same bulk may be as much as five per cent although this is an extreme figure. When one takes into account the fact that the people tested were not experts it is reasonable to suggest that the judgement of a properly trained person comes very near to the limiting errors of the physical measurement.

On reflection, this fact should not surprise us. Certain of the human senses are capable of even more refined sensitivity. It has been claimed that the nose can be used to detect levels of concentration of certain substances which suggest a level of sensitivity higher than that of the spectroscope. It is well known that other substances which may be chemically indistinguishable from one another may be clearly distinguished by the experienced chemist using his nose. This inevitably stresses some of the technical difficulties in the way of developing entirely satisfactory odour standards. Although basic sensitivity is not the sole feature to be considered in the development of satisfactory methods of subjective appraisal, these examples of the high degree of sensitivity of certain of the human senses draws attention to the fact that the *potentialities* of such an approach are considerable. The use to which these potentialities can be put depends very much on systematic training, unsystematised experience being a less efficient means of achieving the same ends.

Other Factors influencing Human Judgement

So far, little or no attention has been given to the many other factors that are likely to affect human judgement. Apart from a few solitary examples, most of the factors to be listed have been 'isolated' in rather academic studies carried out by experimental psychologists. Much information of relevance here is summarised in the chapter entitled 'Principles of Judgement' in the second edition of Guilford's⁷ *Psychometric Methods*. Few attempts have been made to collate the quantitative information about the order of magnitude of the effect of the various factors to be surveyed. The reader will have to take it for granted that detailed references could be supplied to substantiate the facts given below, although few of these references would be concerned with food investigations. The position is in process of being rectified. In particular, reference to American food literature during the last two years indicates that the importance in food assessment of several of the items listed below is in process of being examined (see especially a recent symposium on 'Food Acceptance Testing Methodology'²¹).

The Influence of Irrelevant Characteristics. If several characteristics are given 'scores' or appraised by other techniques, a characteristic that dominates the assessor's judgement may result in other characteristics

which are logically unrelated to the first, being given assessments that are closely related to those of the dominant characteristic. (This effect is well known in the assessment of people. The racing tipster may be assigned qualities of excellence which he does not possess. In the human context this is sometimes referred to as the 'halo effect'.)

The Avoidance of Extreme Categories. Unless deliberate steps are taken to see that a complete range of available categories is used to describe or to codify the characteristics of a number of samples which are being assessed subjectively, then it is very common to find that extreme categories are used very much more rarely than was intended when the set of categories was established.

Serial Effects. These may be of many types. Even when actual comparison standards are provided there is evidence suggesting that judgements are made in relation to the samples assessed in the recent past. How far this interaction extends back is not clearly known. Perhaps this is related to the development of the more permanent types of mental reference standard which may be developed by training or long experience. Other sorts of serial effect include the fact that except in the case of visual inspection (and then only superficially so) it is impossible to make simultaneous comparisons between two samples. This gives rise to a 'time order error' which may be appreciable in its importance. Serial effects may include a number of different phenomena. What is described as 'contrast' may occur when the difference between two samples is accentuated by their proximity (e.g. sourness exaggerated after tasting something sweet). Another effect of relevance here is that of presenting for judgement ascending or descending series of concentrations or intensities of stimuli. For instance, an ascending series of concentrations would lead to an over-estimation of the strength at which the nature of a solution could just be detected. A descending series would yield an under-estimation of the same point. Strictly speaking, these phenomena might equally well be discussed under the later heading of expectation. At any rate, this effect must be taken into account in judging the qualities of foodstuffs.

The 'Odd Sample' in Triangular Tests. For some purposes what are known as 'triangular tests' are used. In these, comparisons are made between three samples, two of which are identical and the third different from the other two. Instances can be given in which the 'odd sample' receives the largest number of correct judgements when the person tested is asked to identify samples having known or prearranged differences. (See the very thorough study of beer tasting by Helm and Trolle.¹⁵)

Marking Labels. Human judgement seems to be influenced even by the marking labels which may be used to identify the samples. For this reason Hening¹⁶ suggested that a combination of letters and numbers which would give the least possible suggestion of systematic

order should be used. No doubt further investigations are required here.

Knowledge of Results. It is extremely important to bear in mind that the results obtained from an investigation are greatly influenced by whether or not those who make the assessments are provided with 'knowledge of results'. Obviously, if the investigator also makes the judgements then he is inevitably in possession of knowledge of results. Not only may this information provide added interest but also it may supply one of the conditions necessary for improvement to take place. Sir Frederic Bartlett has suggested that the old proverb 'Practice makes perfect' should be altered to read 'Practice, the knowledge of the results of which is known, makes perfect'. Thus it is essential to know whether or not knowledge of results has been provided, and this should have been a deliberate decision.

Expectation and Foreknowledge. Human judgement is influenced by expectation and by foreknowledge. For instance, judgements are likely to be easier and more correct if the person making them knows the general type of variation which he is about to meet. In commercial practice, the long-experienced individual will rightly *throw out* anything which seems to be unusual. However, in systematic investigations whose primary function is to increase our knowledge, such actions would not be permissible. Again, whether or not an individual should know what to expect is a matter for conscious and deliberate decision.

Why have Subjective Judgements been under-rated?

Let us now examine some of the possible reasons why the misguided idea got into circulation that in order to be scientific subjective judgements should be eliminated.

1. Most of the early studies concerned with examining 'psychophysical relationships' were carried out in fields where the relevant physical measurements had become highly successful, objective and precise. This state of affairs had been achieved to a considerable extent through the construction of 'artificial environments' which took the form of specially constructed apparatus. The control exerted on the permitted variation of the basic stimuli by the design of the investigations more or less forced a uni-dimensional pattern on the data collected. In fact, it is worth noting that the conditions required for producing the simplest forms of psychological experience are usually highly complex from a physical point of view.

2. In the uni-dimensional situations it has been customary to make the objective variable the independent one because it could be given 'pre-set' values. The subjective variable thus *automatically* becomes the dependent one. It is only recently that statistical techniques have been developed which permit the investigator to take into account the effects of errors in both the variables whose interrelation is being examined. In the past, the effects of such errors have often been assigned, quite

arbitrarily, to the subjective variable. Although this may be of little practical consequence in some instances, in others the interpretation of the data is likely to be biased against the subjective variable unless deliberate steps are taken to see that this does not happen. This problem is of considerable importance when some physical criterion is used to define human sensitivity. According to Bartlett,¹ such a distorting influence does not arise when the independent variable takes prescribed values by deliberate control, but it becomes important when the values 'just happen'. Only in highly academic food studies is it possible to exercise precise control over the relevant objective variables. The possibility of ensuring that samples of foodstuffs take on pre-set values in respect of most of the measurable characteristics is remote. The various 'treatments' or 'processes' may take on accurately pre-set values without the expected results of these processes or treatments being realised in the products concerned. Thus the problem posed here is of direct relevance to food investigation.

3. In many investigations, even those which involve only one objective and one subjective variable, it is easy to commit what at the turn of the century would have been referred to by psychologists as the 'stimulus error'. This involves the assumption that it is the task of the person whose experiences or whose judgements are being examined to appraise what is happening to the physical variable rather than to present certain aspects of what might be described as the resulting sensory qualities. The fallacy of such an assumption may be of negligible practical importance when the 'stimulus situation' and the 'sensory quality' assessed are closely related (e.g. weight and the assessment of weight). However, many sensory qualities do not fall in this category. Clearly, tastes and smells are qualitatively distinct from chemical composition or the basic physical properties involved, although these are in some way the carriers of our experiences. Many characteristics of texture and consistency seem to blend in experience the effects of more than one separately measurable characteristic. For example, Halton and Scott Blair⁸ showed that the springiness of dough as assessed by handling is most closely related to the *ratio* of dough viscosity to dough elasticity. Perhaps this is not the best example that could be found, since some would argue that this ratio defines another physical property, namely the 'relaxation time'. However, as a result of psycho-physical investigations in the same field (i.e. involving flour and dough) the late Professor Katz¹⁷ was led to state that 'the physicist and the psychologist may require different properties, or even a different number of properties, to define the same material in their respective spheres'. These observations are particularly relevant in view of the importance of the consistency and texture of foodstuffs, but they are by no means confined to that limited field.

4. The complex nature of subjective judgements, coupled with their frequent dependence upon more than one separately measurable characteristic, makes any numerical statement of the relation between

subjective and objective information extremely variable and highly sensitive to the context within which the judgement is made. Often a one source of difference is reduced to zero, so other more or less independent sources of difference which were previously of minor importance may continue to give rise to experienced differences. Sometimes such differences may be detected without any realisation that there has been a qualitative change in the origin of these differences. Evidence to support this statement can be drawn from a number of sources, including investigations of the relative firmness of samples of rubber assessed by handling (see Harper,¹¹ p. 27). Since it seems unlikely that useful numerical statements can be obtained which can be generalised far beyond their immediate and specific context, perhaps this has led to an *excessive* emphasis on objective tests. However, the objective testing of many of the 'properties of foodstuffs' involves difficulties similar to those which have just been discussed. The problems of measurement are simple only in those circumstances in which measurement is based upon sound theoretical principles, where control can be exercised within specified limits upon all the relevant conditions of testing, and where the 'properties' being measured are largely unaltered by the processes of testing. Such demands are rarely satisfied by the instrumental tests that can be applied to foodstuffs.

Instrumental Tests versus Human Judgement

It is important to consider the relative merits of the objective and subjective methods available for the appraisal of foodstuffs. The objective methods may include physical, chemical, biochemical and bacteriological tests. As already indicated, at one time it was believed that if only more or better tests could be developed along already accepted lines then subjective appraisal would no longer be necessary. Experience is now leading a number of well-informed food scientists to recognise that certain characteristics of foodstuffs can be appraised effectively *only* by the experiencing individual. This should not really surprise us, for obviously a number of characteristics attributed to foodstuffs are revealed only when the food and the experiencing individual are brought into that close interaction involved in tasting, eating and drinking. We might say that certain food qualities are thus mental abstractions derived from a situation in which the food and the experiencing individual are reciprocally involved. The individual plays the part of the instrument, and yet the interaction is often so complex that the resulting statement may be as much an expression of the 'properties of man' as of the 'properties of the food'. No one could reasonably claim that all food qualities fall in this category, although it may include many characteristics which do *not* belong to the 'preference' group. Surely it is one of the proper functions of applied scientific research in the present field to determine which qualities can be most effectively appraised by subjective methods just as much as it is a legitimate function to develop more and better objective procedures.

It is important to note here that Mulder²⁰ has gone so far as to suggest that one of the reasons why the flavour of (Dutch) cheese has progressively flattened may be accounted for by the fact that *the study of flavour in its own right has largely been neglected in scientific research*. If due attention is paid to the design and the methods of subjective appraisal there is no reason why subjective judgement should not be made just as 'scientific' as certain of the hitherto more widely accepted procedures. It has been suggested that 'quality' means 'fitness for an intended purpose'. Obviously such a definition is too wide for immediate use, but it emphasises the fact that the relative importance of existing subjective and objective methods can only be evaluated when the purpose in mind has been clearly formulated. For some purposes, as for instance in providing information of relevance to chemical processing, instrumental test data may be the most relevant. However, as a guide to the reactions of people some form of subjective appraisal would seem to be indispensable.

A few observations about the word 'objective' will not be out of place here. The ultimate aim of objectivity is to make a statement which is intelligible to others and which may be considered as *the same for all observers*. This does not require us to regard the observer as redundant, nor does it require that the information concerned should be entirely independent of the observer. Objectivity is thus primarily a matter of standardisation of the processes of observation and 'measurement' (using the term 'measurement' here in rather a broad sense). As a result of their work on the quality of white fish, Ehrenberg and Shewan⁶ have shown that satisfactory subjective procedures can be developed, and that through standardisation of these procedures the information collected compares favourably with the usual understanding of the term 'objective'. More recently Ehrenberg⁵ has stressed that even the most respectable forms of physical measurement have a long history which leads back to various forms of unaided sensory judgement in the pre-scientific era. Present-day forms of standardisation and refinement have thus taken centuries of interaction between theory and technique for their development. Hence it is not to be expected that problems of measurement, such as those involved in subjective appraisal, which have been neglected on account of their complexity, will all be solved by a few years of systematic work.

The Examination of Multi-dimensional Relationships

The examination in a systematic and unified manner of the relationship between information collected by means of a number of subjective and objective methods of appraising different aspects of a food product raises many interesting problems. Similar types of relationship are also encountered in our everyday perception of things and events. In this case it is sometimes difficult to believe that such relationships can be examined systematically, for things and events seem to present themselves in our experience in a form which appears completely incapable

of analysis. However, such questions are being tackled by psychologists both from an experimental and a theoretical point of view. It must be admitted that existing techniques for the examination of the relation between experience and its underlying objective aspects have serious limitations. The problems involved may be subdivided as follows:

1. Effectively to record, and possibly to express in numerical terms the salient features of our experiences.
2. Effectively to probe the essential objective characteristics supposed to underlie experience.
3. To develop satisfactory methods whereby these first two items can be brought into relation with each other. This is largely a matter of asking pertinent questions and devising properly planned investigations which are capable of providing simultaneously both objective and subjective information about the same situation.
4. To examine the relationships between the first two items as revealed through systematic analysis and also how to interpret those relationships.

It is important to note that the 'facts' which emerge from such systematic investigations are determined partly by the design of the investigation and partly by the techniques of analysis. Obviously this is a question of high-order complexity. Yet it is one which cannot be ignored. At many points the investigator must make his own deliberate choice between a number of imperfect alternatives. A considerable amount of discussion directly relevant to our immediate problem is to be found in a difficult but extremely important contribution by Brunswik,² which includes a discussion of what he calls the artificial tying or untying of variables. This may arise either from the techniques of investigation or the subsequent methods of analysis. Although Brunswik's important contribution is not directly concerned with the many problems involved in the systematic appraisal of foodstuffs by subjective or other methods, much that he has to say could be translated into such terms.

No further observations will be made here concerning techniques of investigation. However, a brief discussion of the approach to the analysis and interpretation of the relation between a substantial number of variables is necessary. The majority of techniques in use assume that the underlying processes can be adequately represented in terms of simple additive relationships between hypothetical 'elements' which are supposed to remain unchanged from one situation to another. It is extraordinarily difficult to make adequate allowance for the many important serial effects which are known to influence human judgement. Furthermore, the idea of straightforward sequential or chain-like influences itself seems to be a gross over-simplification of the actual nature of the relation between the subjective and the objective aspects of the same situation. In most complex situations such as those which face us here it may be better to conceive of the situation in

terms of a series of reciprocal relationships. Perhaps it is correct to suggest that the model which involves the dependent and the independent variable is appropriate only to the type of situation that is best illustrated by the artificially constrained situations of the classical physical laboratory. Here the apparatus may be designed so as to permit us to adjust one variable at a time while we note the forcibly dependent changes in the other variable. However, this state of affairs cannot really be separated from the artificial environment created by the apparatus itself. When it becomes necessary to consider the simultaneous relationship between a substantial number of variables, some such phrase as 'a network of relationships' is likely to be a more appropriate description of the facts. Exactly what this approach would involve cannot be elaborated here. No doubt valuable analogies might be forthcoming from the field of communication engineering. However, it would be necessary to take into account the fact that in this field most of the objective variables involved have reached a much more finalised and 'respectable' form than can be attributed even to many of the instrumentally measured 'properties' of foodstuffs.

The observations which have just been made presuppose that the main variables of importance are already known. There is much to be said for isolating these variables through a series of patient investigations with strictly limited objectives. If the investigator is not technically well-informed on the particular product it is clear that he will draw upon the expert knowledge of others in the relevant branch of the food industry. Perhaps there are alternative approaches to the piecemeal one which has just been suggested. No doubt a whole symposium might be devoted to considering the purely statistical problems involved in these alternative approaches. One form of 'mass attack' on the isolation of the key variables may be developed through the technique of Factor Analysis. An example of this technique as applied to the study of cheeses will be examined in a later contribution to this journal. Whatever statistical techniques are employed, the 'mass attack' approach itself involves a number of preconceptions, including those which lead to a decision with regard to what should be 'put into' the investigation and the analysis.

Let us now consider certain of the problems raised by the use of systematic factorial designs and their derivatives in investigations involving the subjective appraisal of foodstuffs. Certain of the assumptions underlying these techniques are similar to those referred to above. In the assessment of foodstuffs problems arising from the use of numbers to represent gradations in qualitative data are frequently ignored. These particular problems cannot be examined in detail here, but they fall within the scope of an appeal for the development of statistical techniques whose fundamental models correspond with the *actual* rather than the *assumed* numerical properties of the data examined. However, we have not exhausted all the complications which must be recognised and dealt with in some form or another. Many of the

possible serial or sequential effects affecting human judgement have already been referred to earlier. In some cases these are neither small nor stable. Surely it is necessary to examine whether such influences, whose exact quantitative significance in food investigations is largely unknown, may not play havoc with the assumptions of randomisation of 'treatments' even when this has been carried out according to the classical rules. Randomisation or balanced blocks of treatment seem to be essential if one is to apply existing statistical techniques to establish simultaneously the relative importance of a number of different factors, human and otherwise, in the appraisal of foodstuffs. However, the conditions under which such systematic investigations have to be developed are still more complicated than has been indicated. To the many complications already listed must be added those which are due to the variation between supposedly replicate samples (i.e. samples subjected to the same treatments), to changes in a supposedly constant sample in the course of repeated testing, to differences between individuals and to differences in the same individuals' standards on different occasions. In view of these facts it is not surprising that the writer should propose that if the product and its essential qualities appear to be at all complicated, a piecemeal approach to the development of systematic methods of subjective appraisal is likely to be most economical in effort in the long run. Perhaps a more correct appraisal of the situation may be formulated. A balance must be struck between complex experimental designs, which in other simpler fields of enquiry would provide adequate answers to a substantial number of questions, and designs which are so simple that they fail to reveal the complexities of the issues involved.

REFERENCES

- ¹ BARTLETT, M. S. (1949). 'Fitting a straight line when both variables are subject to error', *Biometrics*, **5**, 207.
- ² BRUNSWIK, E. (1949). 'Systematic and representative design of psychological experiments', *Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press.
- ³ CROCKER, E. C. (1945). *Flavor*. McGraw-Hill, New York.
- ⁴ DAWSON, E. H., and HARRIS, B. L. (1951). *Sensory Methods for Measuring Differences in Food Quality*. Agric. Inf. Bull. No. 34, US Department of Agriculture, Washington, DC.
- ⁵ EHRENBURG, A. S. C. (1955). 'Measurement and mathematics in psychology', *Brit. J. Psychol.*, **46**, 20.
- ⁶ EHRENBURG, A. S. C., and SHEWAN, J. M. (1953). 'The objective approach to sensory tests on foods', *J. Sci. Food Agric.*, **4**, 482.
- ⁷ GUILFORD, J. P. (1954). *Psychometric Methods* (2nd Ed.). McGraw-Hill, New York.
- ⁸ HALTON, P., and SCOTT BLAIR, G. W. (1936). 'A study of some physical properties of flour doughs in relation to their breadmaking qualities', *J. Phys. Chem.*, **40**, 561.
- ⁹ HARPER, R. (1949). 'Food grading and its study', *Food*, **18**, 207.
- ¹⁰ *Idem*. (1950). 'Assessment of food products', *Food*, **19**, 371.

- ¹ *Idem.* (1952). *Psychological and Psychophysical Studies of Craftsmanship in Dairying*, Brit. J. Psychol. Monog. Supp. No. 28. Cambridge University Press.
- ² *Idem.* (1954a). 'The assessment of foodstuffs', *Dairy Ind.*, **19**, 307.
- ³ *Idem.* (1954b). 'Food facts and food fancies', *Nutrition*, **8**, 160.
- ⁴ HARRISON, S. (1952). 'The use of statistics in food processing', *Proc. of the 6th annual convention Amer. Soc. Qual. Control*, 121.
- ⁵ HELM, E., and TROLLE, B. (1946). 'Selection of a taste panel', *Wallerstein Lab. Comm.*, **9**, 181.
- ⁶ HENING, J. C. (1948). *Flavor Evaluation Procedures*, Tech. Bull. No. 284, New York State Agricultural Experimental Station, Geneva, NY.
- ⁷ KATZ, D. (1938). 'The judgements of test bakers. A psychological study', *Occup. Psychol.*, **12**, 139.
- ⁸ McKENNEL, C. Unpublished data concerning the assessment of quality in wool.
- ⁹ MORSE, R. L. D. (1951). 'Rationale for studies of consumer food preferences', *Advances in Food Research*, **3**, 385.
- ¹⁰ MULDER, H. (1952). 'Taste and flavour forming substances in cheese', *Neth. Milk and Dairy J.*, **6**, 157.
- ¹¹ SYMPOSIUM ON FOOD ACCEPTANCE TESTING METHODOLOGY. (1954). (Sponsored by the Quartermaster Food and Container Institute for the Armed Forces, US Army Quartermaster Corps) National Academy of Sciences—National Research Council.

SALES FORECASTING: AN EVALUATION OF SOME PRACTICAL RESULTS*

DAVID J. LISTON, MBE

The Metal Box Company

Mr Liston describes the procedures for making the forecasts required by his firm for budgetary purposes and shows that in the varying conditions that have obtained since the war the forecasts have been reliable enough to be useful. Sales forecasting, he states, is compounded of art, experience, and technical skill.

Background to Budgeting in the Metal Box Company

The Metal Box Company prepared its first annual sales forecast nearly twenty years ago—only a few years after its formation as a public limited company. The Company can therefore claim to have been one of the pioneers in this country in the development of this form of management control. As a result of these early efforts, the Company was able, from the earliest days of the war, not only to estimate with some accuracy the demands which would be made upon its resources by the war effort, but also to feed into the appropriate government departments much of the information required for the control and allocation of the materials in use by the industry, most of which throughout the war, were extremely scarce. By the end of the war therefore, much skill and experience had already been acquired in forecasting the Company's output in terms both of value and quantity, and in forecasting material consumption, on both a quarterly and an annual basis.

This paper deals with our forecasts in the ten years since the end of the war—a period during which I have been associated with the development and directing of forecasting methods within the Company, and with the compilation and co-ordination of the Company's sales forecasts and budgets. It covers a period of rapid change and growth—a period when the Company's activities, at different times and in different sections of the business, have been limited by availability of materials, shortage of labour, production capacity, and the strength of commercial demand.

As an indication of the change and growth of the Company's activities to which these forecasts relate, it is of note that during this period the Company has acquired or built in the United Kingdom thirteen new factories. It has entered into whole new spheres of activity

* This article is based on a paper given to the Study Section (London Group) of the Royal Statistical Society on 10th November 1954.

While the character of many other sectors of the business has undergone radical change. Its total sales in the period have risen by 224 per cent; its consumption of its major material—tinplate—by 65 per cent; its labour force by 42 per cent. During the years under review there have been three major economic crises in the United Kingdom and, in addition, continual changes in the general economic, industrial, and commercial factors affecting the Company's business. There have also been constant changes and fluctuations in prices, both of its materials and products.

Object of the Paper

The object of this paper is to outline the way in which the Company has developed its forecasting methods to meet these changing conditions, to examine the nature and validity of the data on which the forecasts were based, and to measure the results which it has achieved. The paper has been prepared in the hope that our experience, analysed in some detail, will make a contribution to the discussions which have taken place within the Royal Statistical Society as to whether commercial forecasting is: (a) an exact science in which future events can be accurately predicted from a limited number of causal factors, (b) an art in which the knowledge and practice of techniques is supplemented by personal experience and judgement, or (c) a device which fails so far in its claims either to science or to artifice that, whatever its practical value to the business man, the theoretical background to its compilation has merit little serious attention from the statistician.

Results of the Budget

In submitting the results of our efforts to this examination I should make it clear that success or failure of budgeting methods is not properly assessed solely or even mainly in relation to the closeness of budgeted results to actuals. A budget is conceived within the Metal Box Company as an expression of intention by the management with regard to the future operations of the Company. The extent to which the budget is achieved therefore is first and foremost the measure of the success of the Company's operations in that period. Deviations from the budget, whether positive or negative, need to be interpreted in the first instance in the light of management plans and intentions. However, to the extent that over a period of years actual performance falls in line with the budget, there is inevitably reflected the success of those responsible for compiling the budget in co-ordinating and translating into figures the plans, hopes, and intentions of all individuals and departments responsible for the formulation and execution of top management policy.

In effect, the results, as indicated in the graphs and charts attached to this paper, show that in nearly half of the years under review the company has succeeded in forecasting its total sales twelve months ahead to an accuracy of one per cent. In no single year has an error

been more than ten per cent. Its forecasts of turnover for its three main product groups—accounting for ninety-one per cent of the Company's turnover and involving twenty-seven individual forecasts over a period of nine years—have shown in only five cases an error greater than ten per cent. Of fifty forecasts of production for individual products in a single year (1953–54) half the forecasts (representing some seventy-five per cent of production) showed an error of twenty-five per cent or less.

Metal Box Budgeting Procedure

The complete series of budgets prepared by our Company each year includes budgets of sales and production, Head Office expenditure, trading results, and capital expenditure. Additional forecasts are prepared on a quarterly basis and in a rather different way, for longer periods ahead, up to five years. This paper deals in detail only with the annual budget of sales and production. It is concerned with:

- (a) The Statistical Background to the Budget
- (b) The Underlying Economic Factors
- (c) Commercial Intelligence
- (d) Constituents of the Budget
- (e) Budget Co-ordination
- (f) Inherent Errors

The Metal Box Company is primarily engaged in the manufacture of containers from tinsplate and aluminium of a capacity up to five gallons, but it also manufactures kitchen hardware, caps, cartons, composites, drums, and bags of paper and plastic film. In addition, it operates a tinsplate mill, a polythene extruding and converting plant, and a number of factories manufacturing can-making machinery. The total sales of the UK Company, to which the forecasts analysed in this paper relate, amounted to £33 million in the financial year 1953–54.

The Company prepares every year budgets of production (in numbers of containers), of invoiced sales (in value), and of material consumption (in boxes of tinsplate, tons of board, etc.). It may be noted here that the overwhelming proportion of the Company's products are despatched immediately after production and there is little or no accumulation or depletion of finished goods stocks. These budgets are prepared in detail for over three hundred items of production, broken down into budgets for each of the Company's thirty-five factories. The method by which budgets of total sales are translated into individual budgets for each of the individual factories makes a distinct, though interesting, study and is not dealt with in this paper.

(a) *The Statistical Background.* The main stream of the Company's statistics, used both for the preparation of budgets and for the comparison of budgets with actual performance, is based on a system of coding which is used both for orders and for invoiced sales. The codes

ver, amongst other things, (i) manufacturing unit, (ii) customer count numbers, (iii) sales area of customers, (iv) 'basic type' and ecification code of container (six digits in all), (v) trade-use code, i.e. code of the contents to be packed in the container, embracing ninety-nine classifications, and (vi) 'outlet', i.e. government departments, rect export, indirect export, and home trade.

By these codes an analysis is obtained of (i) units of production, (ii) value of invoiced sales and orders, and (iii) consumption of material; and this analysis of past performance in terms of production, sales value, and material consumption, forms the statistical background of the budget as far as internal statistics are concerned.

In addition, we use a variety of external statistics, e.g. production and imports of the main items of canned foods, the Board of Trade index of paint production, statistics of confectionery consumption. In our particular forecasting problems we are chiefly concerned with trends in the trades which pack their products in our containers, and these trades differ widely as regards the extent to which they are documented statistically. We are fortunate that our chief raw material tinplate (together with its own constituents, tin and steel)—is very well covered statistically both on a national and an international basis. We wish this was equally so for some of our other raw materials.

(b) *The Underlying Economic Factors.* Because the Company serves a wide range of industries making consumer goods (particularly in the food trade), its forecast must contain implicitly a number of assumptions about the trend of personal expenditure and the general level of industrial production. In particular, we are concerned with the trends in demand, supply, and price of each of the products packed in our containers—with, for instance, the effect of the government's agricultural policy on the canning of milk or meat; with the relation between house-building and the consumption of paint; with future trends in the price of coffee; with the effect, say, of import restrictions in Commonwealth countries on UK exports of talcum powder, canned fish, and other consumer goods; and with very many other similar problems of economic or commercial significance. We are concerned, too, with the supply, demand, and price, not only of our own raw materials, but also of competing packaging materials such as plastic and glass. A study of the movements of the prices of tin, paper, and board in the last few years (see Figs. 4 and 5) will illustrate the magnitude of the price fluctuations with which we have been confronted.

(c) *Commercial Intelligence.* Our study of economic and commercial data, both national and international, is supplemented by the most careful collection and analysis of commercial information arising out of the company's own operations. For instance, we make use of our own Sales Department to obtain our customers' views on their probable future demands. We analyse and interpret this information in the light of our general economic premises and of our own assessment of

the likely trends in our customers' trade. Our Market Research Section also maintains contact with all levels of the distributive trades thus giving us first-hand knowledge of consumer demand for the products packed in our containers. The basic information for our budget is therefore derived from a network of commercial and economic intelligence which, if not unique, is probably as comprehensive as it is practicable to operate within an industrial company.

(d) *Constituents of the Budget.* On the basis of both internal and external statistics of our basic economic assumptions, of our customers' opinions, of our own market research analyses, and of assumptions we make of the likely trend of our own selling prices, we prepare our estimate of demand, in terms of *numbers* of containers, analysed by 'basic type'. This estimate of demand may then be amended in the light of one or more of the following considerations: (i) availability of materials, (ii) availability of manufacturing capacity, and (iii) availability of labour. It is then expressed in terms of: (i) units of production, (ii) sales value at current prices, and (iii) material usage. At this stage we use the estimate made of the probable trend of our own selling prices in the coming year, and adjustments are made accordingly to the sales value figures.

(e) *Budget Co-ordination.* The compilation of the sales budget involves the managers of every individual factory or group of factories; all regional and product sales managers, and most of the heads of Head Office functional and service departments. Final approval, of course, lies with the top management, to whom the complete budget is submitted for confirmation or amendment.

The process of drafting and co-ordinating is therefore complex. The main responsibility lies with the Information and Statistics Division. The work commences in January (for a financial year commencing 1st April) and takes rather more than two months to complete. Background statistics are prepared in considerable detail and circulated, as appropriate, to all concerned with the budget discussions. There follows a long series of personal visits, consultations, and meetings. Discussion is thorough and often prolonged. The aim is to arrive at every stage at budget figures which will not only be free from purely statistical errors, but also command the confidence and support of those very individuals who will be concerned with their achievement. It is perhaps in this process of negotiation—a process which is often as delicate as it is intricate—that lies one of the main contributions of our Company to the development of budgeting techniques.

(f) *Inherent Errors.* Errors in our forecasts may arise from the following five factors.

(i) CODING. The majority of our codes depend on information readily verifiable within the Company, but we have two codes where

depend entirely on information given to us voluntarily by our customers. Our 'trade-use' code is particularly susceptible to error in the case of plain cans which are subsequently labelled. Thus a canner may, in all good faith, say at the time of placing his order that he intends to use a can for golden plums. If the plum crop fails he may in fact use it to can soup. Similarly, at the time of placing his order a customer may genuinely intend to use a container for export, but when he comes to sell the filled container his export market may be closed and he may be forced to sell it on the home market.

(ii) CONVERSION RATIOS. I have indicated earlier that in its initial stages the forecast is made in terms of *numbers* of different types of containers and that these quantity figures are converted subsequently to figures of sales value and material usage. It is relatively easy, if somewhat tedious, to maintain for this purpose a series of conversion factors based on past performance; but these ratios do, in fact, vary from year to year. An obvious cause of this variation is an increase in selling prices. Other causes are perhaps not so obvious but may be equally significant. The value/quantity ratio, for instance, will alter as containers become more highly decorated. Similarly, the trends towards the use of smaller sizes of can have affected considerably the ratios of both value/quantity and material usage/quantity.

(iii) UNPREDICTABLE NATURAL CAUSES. As manufacturers of containers for food preservation, we find that our results are influenced considerably by natural factors such as the weather affecting agriculture and horticulture. Thus the rainfall in spring affects the milk available for canning; late frosts affect the plum crop; a mild autumn (as in 1953), in addition to reducing demand for soup, increases the supplies of fresh vegetables and reduces the demand for the canned products. The effects of such climatic factors cannot be predicted over the period of the Company's annual budget.

(iv) STOCK CHANGES. I have indicated that, in general, we carry little or no stocks of finished goods, but our containers may remain empty or filled for long periods in our customers' premises and for longer periods in the channels of distribution. Changes in stock levels of manufacturers or distributors can affect demands on us very considerably, and we welcome any attempts by government departments or private concerns to measure changes in stock levels as a matter of routine.

(v) PRICES. I shall refer later to graphs illustrating the variations in prices of our raw materials, which we have experienced since the war. Variations in prices of tinplate, aluminium, paper, and board may substantially affect the accuracy of our forecasts when they are reflected in selling prices. Selling prices are, of course, also affected by changes in wage rates; and in our industry changes in transport rates are another important factor.

Analysis of the Results 1945-46 to 1953-54

It is now possible to examine the results of our budgeting in the light of the information on which they are based and the methods used in their compilation and co-ordination.

Fig. 1 sets out the percentage deviations of total Company sales from the budget for the years 1945-46 to 1953-54. It will be seen that in four of the years the deviation did not exceed 1.2 per cent. The maximum positive deviation over the whole period (in 1949-50) was 6.5 per cent; the maximum negative deviation (in 1952-53) was 9.6 per cent. The nine years show a mean deviation of 3.6 per cent and a median of 4.4 per cent.



FIG. 1. Percentage deviations of the total sales of the Company from budgets, 1945-46 to 1953-54.

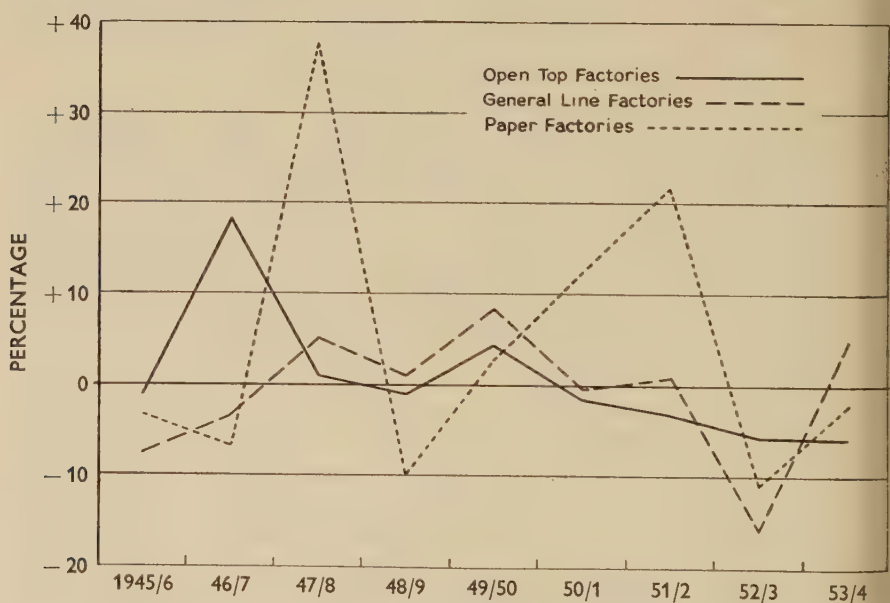


FIG. 2. Percentage deviations of the sales of the main product-groups from budgets, 1945-46 to 1953-54.

Fig. 2 shows the percentage deviation of the sales of the main product-groups of the Company. It will be seen that one group—the Paper Products group—for several years showed exceptionally large deviations. In part this was due in the early post-war years to an inexperience in budgeting sales for this particular side of the business. In one year, however (1951-52), the results showed a 20 per cent

deviation; the price of both paper and board rose very steeply owing to the Korean crisis. Reference to Fig. 4 shows that the price of the paper actually rose 125 per cent within twelve months. It is clear that while in our budget we allowed for some price increase, the full extent of the rise was outside our predictions, as indeed it might well be in the exceptional circumstances then prevailing. It is perhaps significant that in the year of recession (1952-53), while we correctly anticipated the general trend, our results throughout all the groups were rather below the budget.

Fig. 3 gives the actual annual sales of the Company for the past nine years. The general trend of sales has been consistently up, although the extent of the increase has varied considerably from year to year.

Fig. 4 and 5 show the marked fluctuations in prices of some of the main cost-constituents of our products. In many cases material costs are sixty to eighty per cent of the whole, so that price fluctuations in material must have a most marked effect on the evaluation of our output.

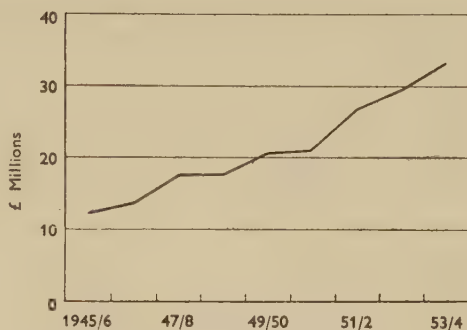


FIG. 3. Total invoiced sales for the Company, 1945-46 to 1953-54.

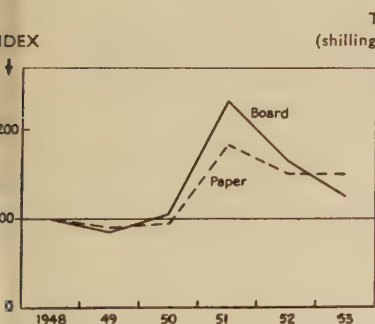


FIG. 4. Indices of average prices of paper and board, 1948-53.

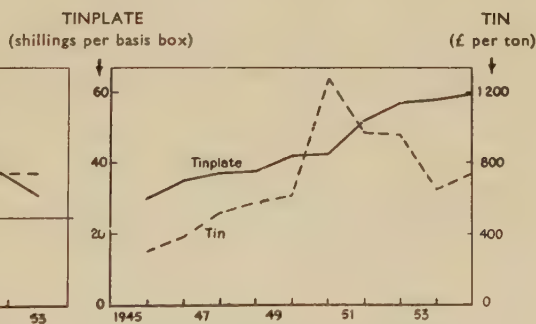


FIG. 5. Tin and tinplate prices at the end of each year, 1945-54. The 'basis box' is the standard unit of area in the tinplate industry; it is defined as 31 360 sq. in.

Fig. 6(a), (b), and (c) illustrates the importance of this factor in budgeting our sales turnover. Our units of production vary enormously in size and character, so that we normally find it of little use to express quantity in a simple aggregate of production units. Instead, we have developed a series of ratios which permit us to express quantity in terms of material consumption. These graphs, therefore, show the wide

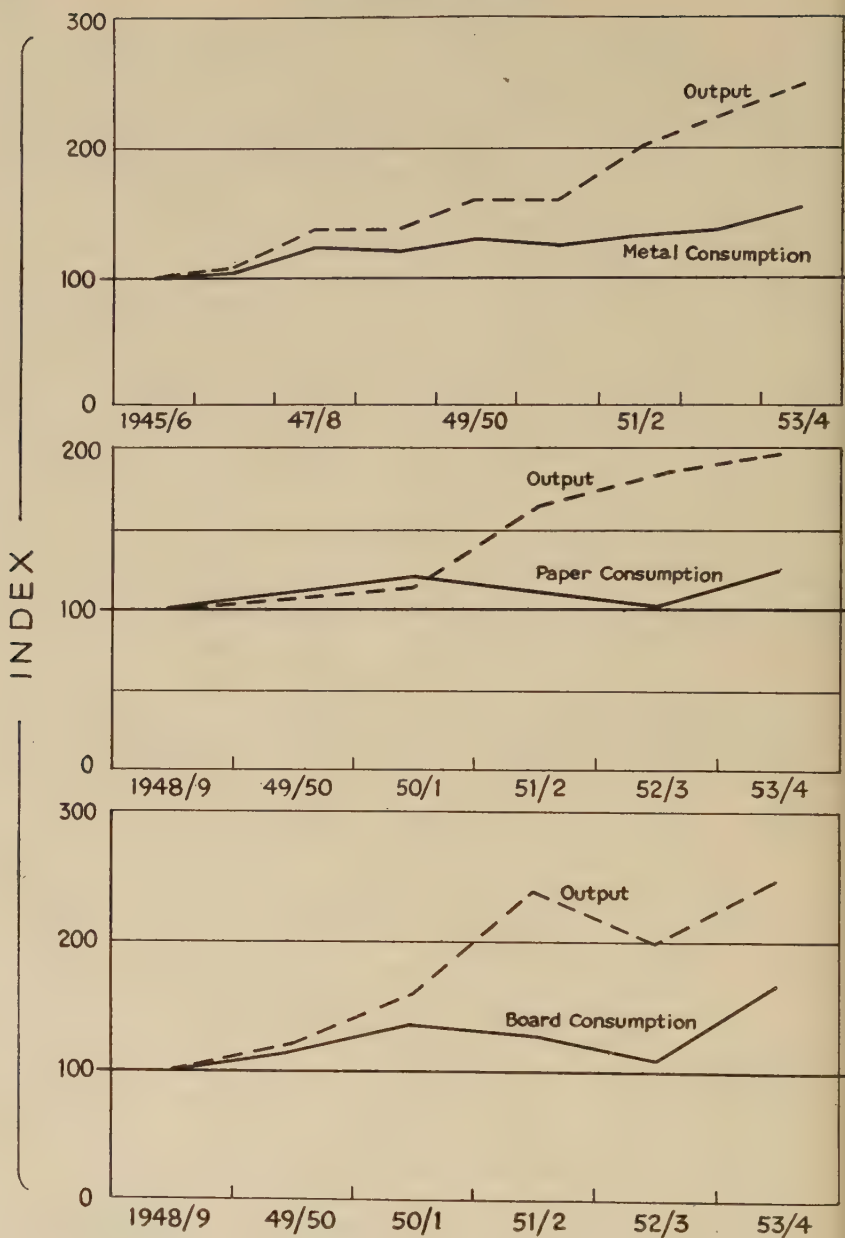


FIG. 6. Indices of the value of output of metal, paper, and board products, and of the volume of consumption of the corresponding raw materials.

divergence between our quantity trend (expressed in this way) and our value trend for three of the main product-groups. It is clear that our budget has to make allowances not only for changing quantities of production but also for changing values per unit of production.

Figs. 7 and 8 illustrate how important it is in our budgets to allow for varying demands of particular industries and for fluctuations in the

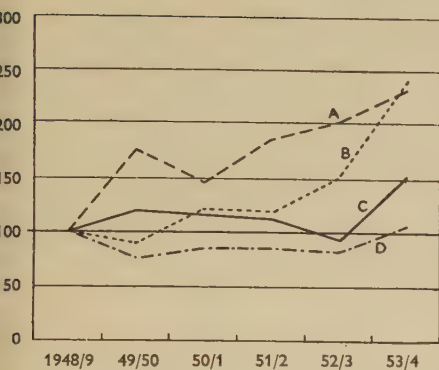


FIG. 7. Indices of despatches of metal containers for industries A, B, C, and D (1948-49 = 100).

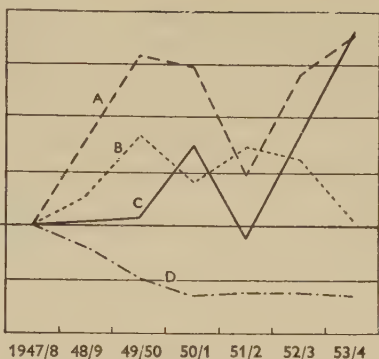


FIG. 8. Indices of despatches of cans for packs A, B, C, and D (1947-48 = 100).

of individual packs. In Fig. 7, for instance, the demand from industry B has nearly doubled itself in five years; industry D, on the other hand, actually took less containers in 1952-53 than it did for the years previously. Individual packs are, of course, very subject to seasonal and other special factors and rarely show a consistent trend.

In Fig. 8, despatches of pack C nearly trebled within a single year, rising in 1953-54 to well over a hundred million cans. Pack B, after one or two promising years, was in 1953-54 once again in the doldrums. Pack A has had its vicissitudes, but is again hitting new peaks. Pack D, on the other hand, seems consistently depressed. Each of these individual packs needs the most careful assessment in each year if the budget is to bear any relation to actual sales.

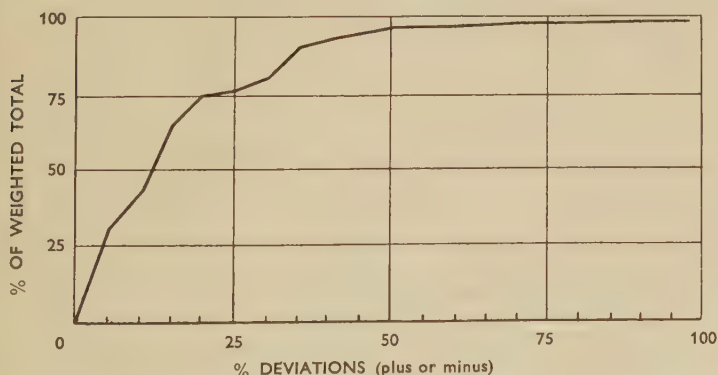


Fig. 9. Cumulative distribution of the weighted deviations of actual from budgeted quantities 1943-54. The cumulative total amounts to 98.4 per cent. The remaining 1.6 per cent is accounted for by three items with positive deviations of more than 100 per cent.

Finally, Fig. 9 shows what happens to some of the more detailed estimates contained in the budget. There are in all possibly a thousand different items for which individual forecasts are made. It is clear that with each refinement of detail there comes an increased liability to

error. For all that, in this particular year (1953-54) well over three-quarters of the Company's output was budgeted in detail with an error not exceeding twenty-five per cent. Many of the divergencies were, of course, the result of deliberate management decision and were, in fact, inevitable; for if no such departures are permitted to meet changing conditions during the year, the management loses its flexibility. If that happens the budget, instead of being (as it should be) a supple and delicate instrument of control, is in danger of becoming a strait-jacket.

Conclusion

The conclusion drawn from the paper is that the experience of our Company is such that it may fairly be claimed that sales forecasting, as applied to our own efforts, has achieved a degree of reliability which removes it well outside the field of mere guesswork. Exact science it can most probably never be. We use no econometric methods, nor, indeed, have we succeeded so far in establishing correlations or multiple regressions upon which to base at least some portion of our forecasts.

On the other hand, we do use the scientific method in the compilation of our budget. We first collect, collate, inspect, and analyse our data—both statistical and otherwise. We then formulate our premises, from which we attempt to derive logical conclusions. Finally, we present results in a way as free of errors in fact or bias as we can contrive.

Compounded as it must be of art, experience, and technical skill, sales forecasting can claim for itself a special place in management theory, meriting attention not only from managers and executives but also from statisticians and others who seek to explore the theoretical background to modern management practice.

LEVEL OF QUALITY FOR MINIMUM COST OF MANUFACTURE TO A SPECIFICATION

JOHN H. HEPBURN

Imperial Chemical Industries Ltd., Nobel Division, Stevenston

If loaves of bread are required to have at least a certain specified weight, the average weight must be greater than this in order to avoid an undue proportion of loaves that are underweight. The higher the average weight the greater is the cost of manufacture but the smaller is the cost due to the occurrence of underweight loaves. What average weight gives the lowest total cost? Mr Hepburn's article is about this type of problem.

Introduction

Measurements of physical and chemical characteristics, such as density or viscosity, of samples of a manufactured product usually exhibit some variation about a mean value. The frequency distribution of the characteristic can be determined and the probability of extreme values calculated. If it is desired to control the characteristic within certain limits its mean value is so chosen that values outside these limits are unlikely to occur solely because of random variations.

When the frequency distribution is normal and both upper and lower specifications have to be met, the best value of the mean is usually midway between the specification limits. There is then least chance of some of the product being rejected because of values occurring outside these limits. It is also clear that the proportion of rejected product may be reduced by reducing the standard deviation of the measured characteristic.

On the other hand, some products are required to meet only an upper or lower specification, and by suitably choosing the level of the mean value the proportion of rejected product can be made as small as desired. Often, however, the cost of the product is dependent on the mean of the measured characteristic, and in order to obtain minimum cost of manufacture it is desirable to work as closely as possible to specification. For example, if the weight of the contents of a package is to be not less than a given nominal minimum and if the selling price is fixed and based on that nominal weight, it is obviously wasteful to give away more than is barely necessary to ensure that all packages just meet the specified weight. However, the closer the mean to specification the greater will be the proportion of material rejected; and if the rejected material is destroyed or if the cost of reprocessing is high, it is very important to find the optimum value of the mean which gives the most economical product.

As far as working to a lower specification is concerned the problem is of wide interest and will be particularly so if the recommendations of the Hodgson Report¹ are accepted and it becomes illegal to sell any product containing less than the amount purported. The solution usually adopted is to work to a mean value at about three or four standard deviations above specification and to check by means of a sample that the mean and standard deviation are within prearranged limits. However, if only a sample is tested it can never be possible to ensure that all items are within specification, so that there is always bound to be some risk of a consumer receiving an item which should have been rejected. When no consumer's risk can be tolerated 100 per cent inspection must be carried out. Alternatively, if savings can be effected by working to a lower mean value and rejecting items below specification it may be found that these savings more than pay for the cost of 100 per cent inspection. This paper is relevant only to those cases where for one reason or another all the product is tested.

When working to an upper specification a similar problem may be found. For example, if the moisture content of a product is to be kept below some value and the greater the percentage of moisture the cheaper the product, the problem is to determine at what level of moisture content to work. Again, if the percentage impurity in a product is not to exceed a specified maximum value it would be unnecessarily costly to introduce elaborate processing to remove the impurity altogether, provided that the circumstances did not render this desirable or necessary. Such problems are not considered separately in this paper, as the solutions may be easily obtained from those given for the cases in which the lower limit is specified.

Product Outside Specification Destroyed

In some processes the product falling outside the specification is destroyed, so that the cost of the material actually used is increased; e.g. if it costs $\pounds A$ to make unit product and a proportion P is destroyed the cost of manufacture of unit acceptable product is then $\pounds A/(1 - P)$. If A is constant this cost is a minimum when P is zero, but if A depends on some mean value characteristic of the material it may be possible to obtain minimum cost at a mean value which would necessitate a proportion of the product being destroyed.

Assume that the cost of unit quantity of the product is made up of a constant cost and a cost which is proportional to the mean of a characteristic variable, so that the total cost of manufacture of unit quantity is of the form

$$C = a + bm \quad \dots (1)$$

where m is the mean level of the characteristic variable x ,

b is the marginal cost per unit of x ,

and a is a constant cost independent of x .

Suppose x is normally distributed with standard deviation s_x . Let $(m - x)/s_x = t$ and $(m - L)/s_x = u$, where L is the lower specification limit. The proportion P of the product outside specification is therefore equal to the probability that t is greater than u . Denote this by

$$P = P(u) \quad \dots (2)$$

If all rejected material is destroyed and c is the cost of destroying unit quantity, the actual cost of unit quantity of the accepted product becomes

$$K = \frac{a + bm + cP}{1 - P} \quad \dots (3)$$

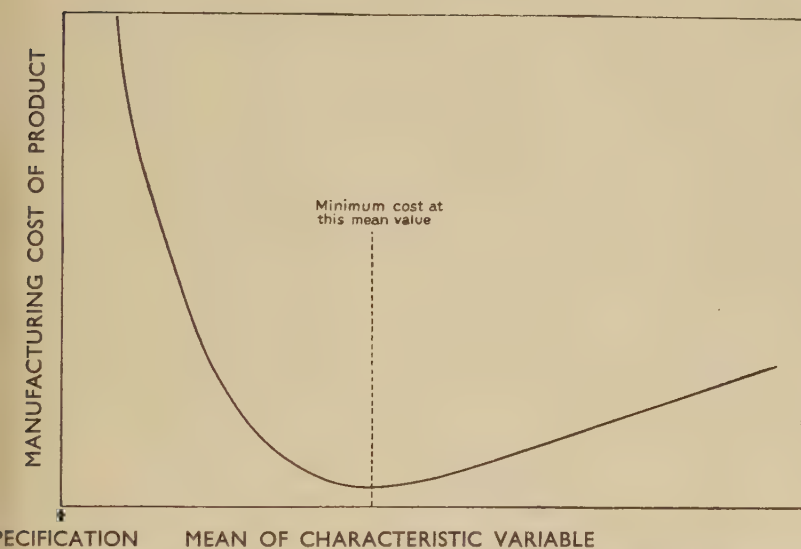


FIG. 1. Usual type of curve obtained when the cost of manufacture of the product is plotted against the mean value of a characteristic variable on which the manufacturing cost is dependent.

When K is plotted against m the curve so obtained may not have a minimum under certain conditions, but on the assumptions made here, namely those of normality and the admissibility only of positive values of constants and variables, it will generally have the shape of the curve drawn in Fig. 1. It is not always convenient to draw this curve in order to find the value of m which gives minimum cost, but the minimum may be found by differentiating K with respect to m and equating to zero. This yields

$$(1 - P)(b + cP') + (a + bm + cP)P' = 0$$

and since $P' = -z/s_x$, where z is the ordinate of the normal curve at u , we have

$$\frac{1}{s_x} \left(\frac{a + c}{b} + L \right) + u = \frac{1 - P(u)}{z} = f(u) \quad (\text{say}) \quad \dots (4)$$

The right-hand side of (4) is a function of u only, and so the plot of this against u may be drawn once for all. The left-hand side can be represented by a straight line of unit gradient, so that a family of lines of unit gradient but differing intercepts may be drawn (Fig. 2). Once a , b , c , L , and s_x are known the optimum value of m can be immediately found from the intersection of the appropriate straight line with the curve representing $f(u)$.

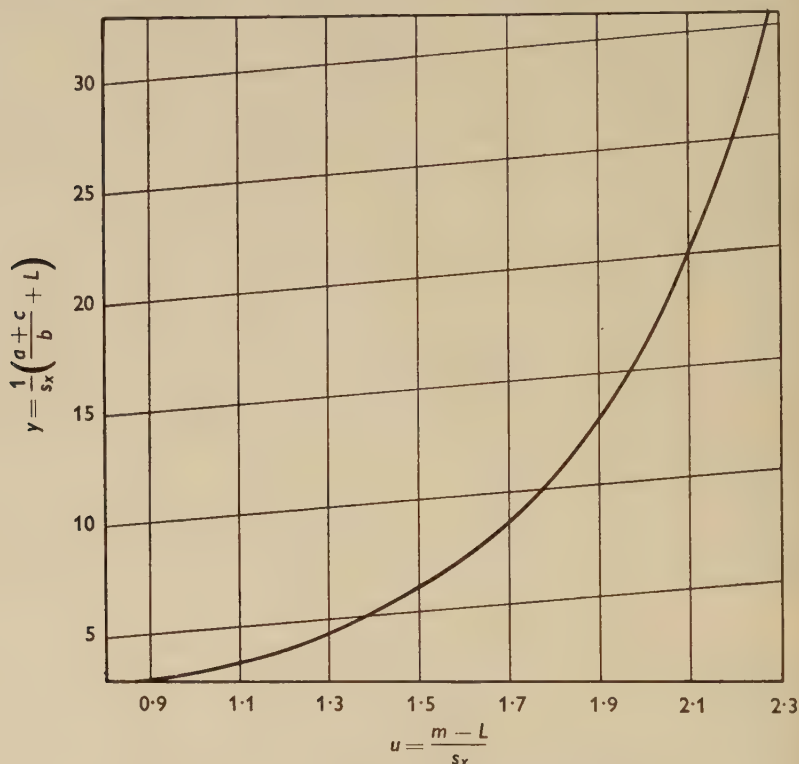


FIG. 2. Curve for determining the mean of the characteristic variable to give the minimum cost of manufacture when rejected product is destroyed.

It can easily be shown that the slope of $f(u)$ is greater than unity for all positive and less than unity for all negative values of u , so that a straight line of unit gradient will intersect $f(u)$ either twice, at a positive and a negative value of u , or not at all. The positive and negative values of u refer to the respective values of m for minimum and maximum cost. When $u = 0$, $f(u) = 1.253$, so that if the line

$$y = \frac{1}{s_x} \left(\frac{a + c}{b} + L \right) + u$$

intersects the y -axis at a value less than 1.253 it will not meet the curve $y = f(u)$; that is, there will be no solution to (4) if

$$\frac{1}{s_x} \left(\frac{a + c}{b} + L \right) < 1.253$$

In practice the condition for no solution will not arise. Constants and variables cannot be negative; further, we have assumed a normal distribution, and for this assumption to be valid m/s_x must be greater than 3, even when u is small, and hence in the limit L/s_x must be greater than 3. If $a = c = 0$, the intercept of the line of unit gradient on the y -axis is L/s_x , and so the intercept must always be greater than 3. The line of unit gradient making an intercept of 3 on the y -axis meets the curve $y = f(u)$ at a value of u of 0.9, and so the optimum value of m must always be greater than $L + 0.9s_x$.

Example

The figures quoted here are quite fictitious, although a similar problem does exist in the bread-making industry.

Loaves of bread must be at least 28 ounces in weight. From constant dough weights of 30 ounces the weights of baked loaves are normally distributed about a mean weight of 28 ounces with a standard deviation of 1 ounce. Also, the weights of dough from a dough divider are normally distributed about the nominal weight with a standard deviation of 0.75 ounce. The cost of dough is 0.2 penny per ounce and the cost of baking and overheads is 0.5 penny per loaf. The cost of destroying rejected loaves is negligible. What is the optimum weight of dough from a dough divider to give a loaf of minimum cost?

There are two sources of variation: the dough divider and the oven, so that the total variation will be the sum of the variations due to these sources, i.e.

$$s_x^2 = 1^2 + 0.75^2 = 1.56$$

Also $a = 0.5$, $b = 0.2$, $c = 0$, $L = 30$.

$$\therefore y = \frac{1}{s_x} \left(\frac{a + c}{b} + L \right) = 26$$

From Fig. 2, u is found to be 2.2. Therefore $m = 30 + (2.2)(1.25) = 32.75$ ounces.

Suppose a market were available for underweight loaves at 4 pence each. The cost would then be of the form $(a + bm - cP)/(1 - P)$, but the previous solution can be used provided c is put equal to -4 . Then

$$\frac{1}{s_x} \left(\frac{a + c}{b} + L \right) = 10$$

and reference to Fig. 2 gives $u = 1.73$. Therefore $m = 32.16$ ounces.

Product Outside Specification sold at Cheaper Price

It will be noted that it may be possible in cases such as the one just considered, where a market is available for material below specification and c (the selling price of the rejected material) is high, that

$$\frac{a - c}{b} + L < 1.253s_x$$

This has already been seen to be the condition for no solution to (4). It is equivalent to $c > a + bL - 1.253bs_x$. That is, for no solution, the selling price of material below specification must be fairly near the cost of manufacture of material to specification. However, if it is important to meet the specification there must be some incentive to do so. This incentive is usually provided by the fact that the selling price of material below specification is so low that production of such material is unprofitable. Cases of no solution should therefore not arise in practice.

Error of Measurement

No account has been taken so far of the error of the measuring apparatus. Owing to this error the indicated level of the characteristic may not be the true one, so that some satisfactory items may be rejected and other unsatisfactory ones accepted. This will not be very serious if the measuring error is small, for it will mean only that there will be some doubt whether material very close to specification will be accepted or rejected.

In the type of problem considered here, the consumer's risk, i.e. the chance of passing as satisfactory material which should be rejected, should be as near to zero as possible. From intuitive reasoning it is obvious that the greater the measuring error the greater will be the consumer's risk, but an exact mathematical relationship can be easily derived.

The measuring error will most likely be normally distributed about zero mean with a certain standard deviation, s_e say. It has already been assumed that the characteristic is normally distributed about a mean m with a standard deviation of s_x . Hence the probability of a characteristic value less than L and an indicated value greater than L , i.e. the probability of an unacceptable unit being accepted, can be expressed in the form

$$M(u, -ru, -r)$$

where $r = s_x/(s_x^2 + s_e^2)^{1/2}$ and $M(h, k, r)$ is the bivariate normal function which has been tabulated by Pearson.² Further,

$$M(h, -k, -r) = 1 - P(h) - M(h, k, r)$$

and so the consumer's risk may be calculated. Similarly, the probability of rejecting an acceptable item, i.e. the producer's risk, is

$$M(-u, ru, -r) = 1 - P(ru) - M(u, ru, r)$$

Consumer's and producer's risks have been calculated for different mean values and for different values of the ratio s_e/s_x and are presented as families of curves in Figs. 3 and 4. It is clear from Fig. 3 that if the mean is more than two standard deviations above specification there is little consumer's risk, provided that the standard deviation of the measurement error is less than that of the production distribution.

If in any particular case the consumer's risk is not small enough to be ignored, it may be made as small as desirable by an appropriate increase in the test specification. Thus, if the product specification is $m - u s_x$ the test specification would be set at $m - u s_x + v s_e$. The

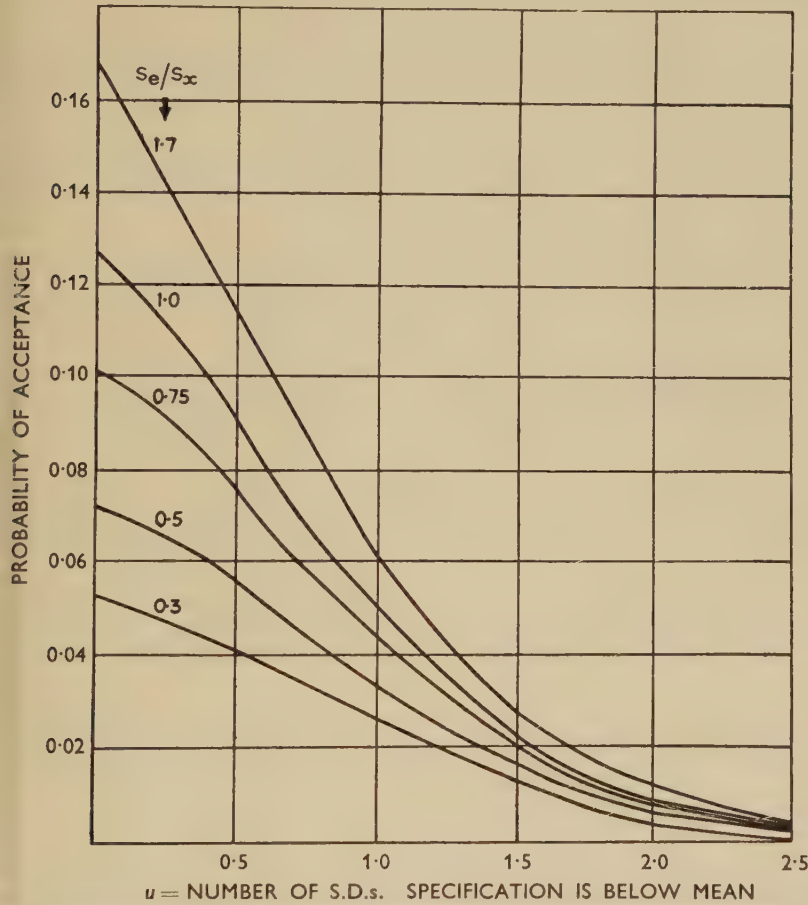


Fig. 3. Consumer's risk for varying mean values and for different levels of the ratio s_e/s_x .

Higher the value of v the lower will be the consumer's risk. Eagle³ has considered this problem when both upper and lower specifications have to be met. Grubbs and Coon⁴ have extended Eagle's solution to cover cases when consumer's and producer's risks are equal, the sum of the consumer's and producer's risks is a minimum, and the cost of making a wrong decision is minimised.

Product Outside Specification Recovered

In most cases the product falling outside specification is not destroyed. Either an adjustment is made to the rejected items to bring them within specification or they are stripped and the constituent materials recovered and reprocessed.

Assume, as before, that x is the characteristic variable, m its mean, s_x its standard deviation, and L the specification level. Let M be the mean value of x in the recovered proportion. Again, assume that the cost of manufacture of unit quantity of the product is made up of a

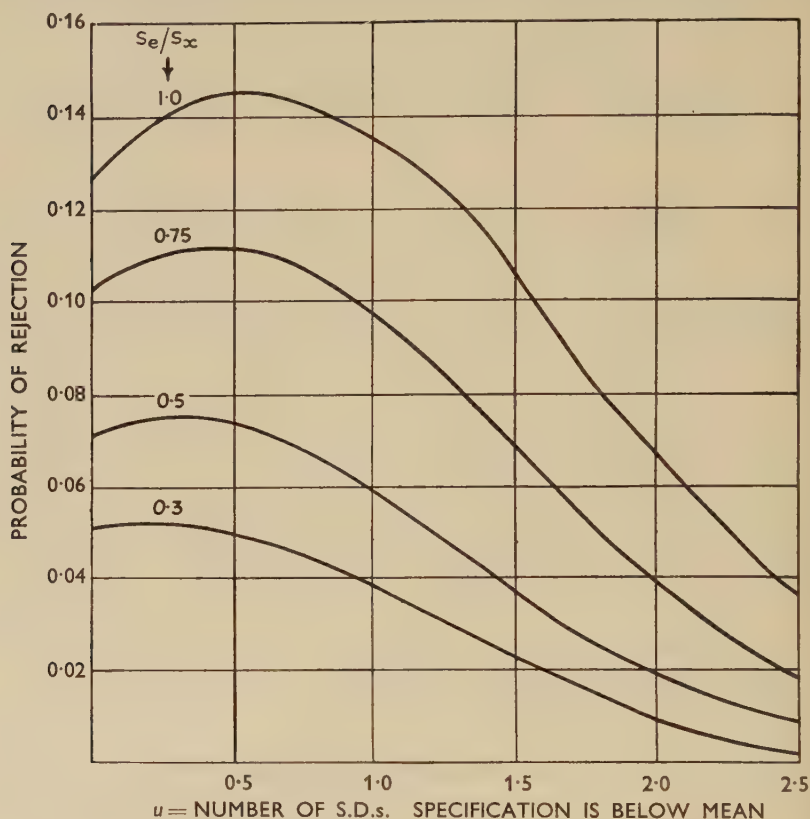


FIG. 4. Producer's risk for varying mean values and for different levels of the ratio s_e/s_x .

constant cost and a cost which is proportional to the mean of x . If a proportion P is rejected the actual cost of unit quantity of the accepted product is

$$K = \frac{a + bm + cP - bMP}{1 - P} \quad \dots (6)$$

where c now is the cost of reprocessing unit quantity.

If x , m , and L are measured in units of s_x so that, in effect, $s_x = 1$, then

$$M = mP(u) - z$$

and the operation of putting $dK/dm = 0$ yields

$$\left\{ u + \frac{z}{1 - P(u)} - \frac{1 - P^2(u)}{z} \right\} + L\{1 - P(u)\} + \frac{w}{1 - P(u)} = 0 \quad \dots (7)$$

where $w = (a + c)/b$.

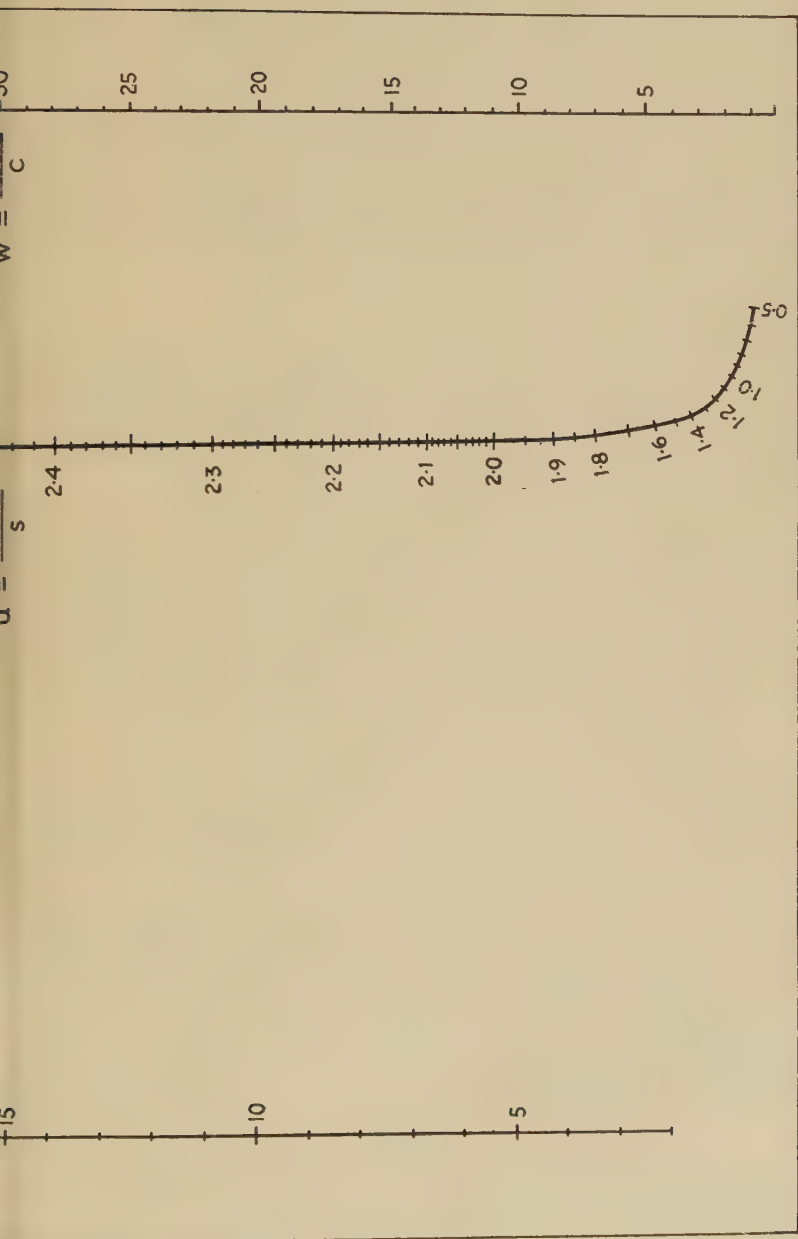


FIG. 5. Nomogram for determining the mean of the characteristic variable to give the minimum cost of manufacture when rejected product is reprocessed.

Since z is a function of u , (7) can be written

$$F(u) + LQ(u) + w/Q(u) = 0$$

or

$$\left| \begin{array}{ccc} w & 1 & 1 \\ L & 0 & 1 \\ -\frac{F(u)}{1+Q^2(u)} & \frac{1}{1+Q^2(u)} & 1 \end{array} \right| = 0$$

so that a nomogram (Fig. 5) consisting of two straight lines and a curve can be prepared for the solution of (7). Once all constants have been determined, the value of u which makes K a minimum can be read off and the corresponding value of m calculated.

Example

In chocolate manufacture a volume of molten chocolate to give a bar of the correct nominal weight is poured into a mould and chilled. Subsequently the bars are automatically weighed and any underweight rejected. The rejected bars are recovered and reprocessed. For the purpose of this illustration we shall assume certain costs and values which may have been exaggerated.

For bars of nominal weight two ounces the weights are normally distributed about the nominal with a standard deviation of 0.09 ounce. The mean error of the automatic weighing device is zero with a standard deviation of 0.03 ounce. The fixed constant cost in the manufacture of a two-ounce bar is 0.7 penny, the cost of the chocolate is 2.5 pence per ounce, and the cost of reprocessing a rejected bar is 0.8 penny. Hence,

$$a = 0.7, \quad b = 2.5, \quad c = 0.8, \quad w = 0.06$$

$$L = 2, \quad s = (s_x^2 + s_e^2)^{\frac{1}{2}} = 0.095, \quad L/s = 21$$

From Fig. 5 we find that the line joining the points 0.06 on the w scale and 21 on the L/s scale cuts the u scale at a value of 2.1 and so

$$m = 2 + (2.1)(0.09) = 2.2 \text{ ounces}$$

i.e. a volume giving a nominal weight of 2.2 ounces would require to be poured. Moreover $s_e/s_x = 0.33$, and reference to Fig. 3 shows that there is a consumer's risk of about 0.2 per cent due to the error of the automatic weighing device.

The author gratefully acknowledges the advice of his colleagues in the preparation of this paper.

REFERENCES

- ¹ HODGSON, SIR E. H. (Chairman) (1950). *Report of the Committee on Weights and Measures Legislation*. Cmd. 8219, HMSO, London.
- ² PEARSON, K. (1913). *Tables for Statisticians and Biometricians*, Part 2. Cambridge University Press.
- ³ EAGLE, A. R. (1954). 'A method for handling errors in testing and measuring', *Ind. Qual. Control*, **10**, No. 5, 10.
- ⁴ GRUBBS, F. E., and COON, H. J. (1954). 'On setting test limits relative to specification limits', *Ind. Qual. Control*, **10**, No. 5, 15.

MEASURING THE READERSHIP OF NEWSPAPERS AND MAGAZINES

ERIC SHANKLEMAN

Mr Shankleman has written this article around a review of a readership survey, and has compared the results of the survey with those of another one covering roughly the same period. He has also considered the use of such results in advertisement planning.

ADVERTISERS in Great Britain are estimated to have spent just over 199 millions on advertising in 1952.¹ In 1953 total advertising expenditure is estimated to have been some £280 millions.² These are vast sums of money. Yet very little is known, with any degree of precision, as to how effective this expenditure is in creating demand for products advertised.

Observation of the business world indicates that some enterprises have prospered and grown with the minimum of advertising in the normal media of press, posters, radio, cinema, and the like. But, on the other hand, there are far more businesses which believe implicitly in the need for large-scale advertising to maintain and stimulate demand. The need to keep the market informed of the existence of a product, and of its attributes, is one of the tasks of advertising, and in this sense that this is achieved it performs a very useful function. But how much should be spent on this process, and whether an expenditure by an individual advertiser of £60 000 per annum is likely to have much more effect than one of £50 000, are highly pertinent questions which, it is suspected, are often inadequately probed and seldom satisfactorily answered.

This is not to say that the advertiser plans his expenditure entirely 'in the dark'. Considerable research, for example, has been directed towards ascertaining the extent to which different advertising approaches appeal to samples of the public for which they are intended, although the results of these tests have usually not been measured in terms of sales results. Where they have been so measured, the tests have been local rather than national.

Readership and Circulation

More extensive, however, has been the research into the numbers and characteristics of the people reading different newspapers and journals. Well over fifty per cent of the £200 millions and more spent on advertising in each of the two years quoted above was accounted

for by the cost of space bought in the press and in magazines. The aim of such readership research is to indicate to the advertiser how many people of the type likely to be interested in his product may see his advertisement; readership in this sense measures the size of the potential audience.

Regular information on the circulation of newspapers has been available in this country since the founding in 1931 of the Audit Bureau of Circulation. But this is restricted to data on the audited number of copies sold during a given period, and throws little light on the kind of people buying the paper or periodical. Who they are, for example, is not answered, although many individual journals have attempted from time to time to classify circulation figures by region. To answer such questions as 'who?' sample surveys are necessary. The history of these surveys in Great Britain has been set out in a recent study by the Market Research Society of Great Britain.³ The first survey was carried out in 1932; a second followed in 1936, and a third in 1939. Since the end of the war in 1945 there have been many more such surveys; in 1947 no fewer than three specific readership surveys were made. One of these, by Hulton Press Ltd., became the first of a series of annual surveys sponsored by that company.

Sampling Methods

All except one of the surveys carried out since the war have used quota samples, most of them consisting of some 12 000 or more adults, in order to provide reasonably satisfactory bases for estimates of sub-categories and of the readership of less widely known publications. The use of quota sampling enabled the field work involved in these surveys to be carried out less expensively than would have been the case if random sampling designs had been used.

A random sample was, however, adopted in the recent National Readership Survey sponsored by the Institute of Practitioners in Advertising.* The sample, devised in consultation with Professor Maurice G. Kendall, used a four-stage design, and involved the selection of 720 interviewing districts, in each of which 25 interviews with adults over the age of 21, as chosen from electoral rolls, were planned. It had been suspected that the electoral rolls might under-estimate the numbers of persons around the age of 21 in the population, and this deficiency would clearly be reflected in the sample of individuals drawn from the rolls. To overcome this, and at the same time to enable a sample of those aged 16-20 to be drawn, a separate sample of addresses was taken from the electoral rolls. At such addresses all persons aged 16-24 inclusive were interviewed. In the event, the fears of the IPA were substantiated: the electoral lists were shown to understate the numbers aged 21-24 and, to quote the IPA report, 'not only was this

* Institute of Practitioners in Advertising: *The National Readership Survey—1954 Edition*. London, 1954. Pp. 115, 45 tables. £10 10s.

group numerically deficient but it was biased towards the middle and upper classes.'

An admirable feature of the report of this survey is the very detailed description given of the sampling design and interviewing procedure. Failure to include similar information has rightly invoked criticism from other reviewers of similar surveys.⁴ Of particular interest is a table detailing reasons for interviewers' inability to carry out interviews. Of the persons drawn for interview, some 3·8 per cent had died, or had removed to another address before the qualifying date of the electoral register, or were not locatable for other reasons at the time of the proposed interview. In all, some 2 417 people within the sample were found to have moved, but 1 386 of them were followed up at their new address.

The discussion of sampling methods in the IPA report includes other interesting information. One often wonders, for example, to what extent people refuse to be interviewed. The figure given is 3·9 per cent, but this relates not only to those who refused to see the interviewer but to those who failed to complete the interview; it is a pity that separate figures were not quoted, especially in view of the somewhat lengthy interview involved in this survey. Considerable persistence was practised by the investigators; except where a definite refusal was met, or where circumstances made an interview impossible, at least three recalls were made to achieve contact, and more than three when there seemed hope of success. Just over half of the successful interviews needed two or more calls; and over 54 per cent were carried out after 5 p.m. Their persistence was rewarded by the very high response rate achieved, a figure of 85·6 per cent of individuals selected from the electoral register. This is probably at least as high a response rate as has been achieved in this country in a survey using a random sampling design. Clearly, for workers in the market research field who are responsible for sampling design and the organisation and control of interviewing, their experience could be of great interest.

Definitions of Readership

A perplexing feature of the reports of readership studies to the users of the results must be the variations in definition of readership adopted by the different institutions responsible for carrying out surveys. In a series of surveys carried out by the Research Department of Odhams Press⁵ the reader is defined as one who 'reads or looks through' a newspaper on the day prior to the interview. A survey published by the Kemsley Press Ltd. in 1954⁶ defined reading as 'the conscious looking through or reading of more than one page of the specified issue of a paper'—namely, the paper appearing on the previous day. This definition was purposely designed 'to imply more than a casual skim through the headlines and illustrations'. Other surveys have attempted to measure *proved* readership, where evidence has been sought that a particular issue of a paper has been seen and read. In the Hulton

Survey⁷ the reader was asked, for example: 'Which daily morning newspaper did you read yesterday?', a definition of readership which has been adhered to since the series started in 1947. Thus the Hulton figures 'show the proportions and numbers of people who say they have read (in the broadest sense) a particular daily newspaper on an average weekday, a particular Sunday newspaper on an average Sunday, a particular weekly magazine in any one week, and finally, a particular monthly magazine in a period of four weeks'.

In the National Readership Survey sponsored by the IPA the key phrase used in the questionnaire was 'looked at'. Those interviewed were asked firstly: 'Do you happen to have looked at a copy of the — any time over the past six months?' If the answer were 'Yes', the respondent was then asked: 'When was the last time you looked at a copy of it?' If the daily newspaper being considered had been read yesterday, or the Sunday newspaper—to take another example—within the last seven days, the respondent was asked: 'Where and when was it seen?' The IPA report explains that this definition was used in preference to alternatives because it was considered to be 'colloquial enough to be understood with the right shade of meaning for the purpose of the investigation'.

TABLE I

Proportions of adult population reading certain types of publication and average number of publications read per reader

	Readers* per 100 Adult Population		Average Number of Publications 'read' per Reader		$\frac{100 \times \text{Col. (4)}}{\text{Col. (3)}}$
	Hulton (1)	IPA (2)	Hulton (3)	IPA (4)	(5)
<i>Males</i>					
National daily newspapers ..	82.5	84	1.25	1.68	134.4
National Sunday newspapers ..	92.6	92	2.05	2.40	117.1
General weekly magazines ..	76.4	74	2.09	2.00	95.7
<i>Females</i>					
National daily newspapers ..	76.4	78	1.20	1.43	119.2
National Sunday newspapers ..	89.6	89	1.99	2.23	112.1
General weekly magazines ..	69.9	71	1.87	1.83	97.9

* A respondent qualified as a reader if he reported reading any of the publications included in the list of daily newspapers, Sunday newspapers, or weekly magazines. The published information does not readily permit sufficiently similar comparisons to be made in respect of general women's weeklies or women's monthly magazines.

Comparisons of the IPA and Hulton Estimates

For many reasons the reviewer of the IPA Survey is tempted to compare its results with those of the 1954 Hulton Survey. Both were ostensibly designed to measure the same activity—readership—although with different definitions, within a short period of each other. Both used large samples, but in the IPA Survey a random design was used, and in the Hulton Survey a quota design. Table I gives a comparison of some of the results of these two surveys. It will be seen that the estimates of the proportion of adults who claimed to have read at least one of the publications in each of the groups listed in the table are very close in most cases (columns 1 and 2). But much bigger discrepancies, which cannot easily be explained, occur between the average number of national daily and national Sunday newspapers read per reader, and in each case the IPA estimates are higher than those of the Hulton Survey.

It is interesting to note that this result would seem in some ways to be contrary to those obtained in the carefully controlled research by Moser and Stuart into comparisons between quota and random samples.⁸ They state: 'It has sometimes been noticed that on activities such as cinema-going, reading, smoking . . . quota samples tend to get higher percentages than random samples', and suggest that the reason is that quota samples get too educated a cross-section of the population. They go on to say: 'It is clear that in surveys closely related to education—such as Readership Surveys—this danger of bias cannot be ignored and that, if possible, some kind of educational control should be used.'

Differences between the number of papers read per reader are reflected in Table II, which shows the proportions of male adults reading various publications. It is seen that in the case of individual daily newspapers the IPA estimates are consistently higher than those of the Hulton Survey. On the basis of the estimates of standard errors given in the IPA report, it is highly improbable that many of the divergencies from the IPA figures could be explained by chance. In the case of the national Sunday newspapers, IPA estimates of the percentages of adults reading the papers concerned are also higher than the comparative Hulton figures, although the differences tend to be less significant. The estimates are much closer for weekly periodicals, with the Hulton estimates tending to be larger, and the same is true for the two monthly periodicals included in the table.

Some Possible Causes of Differences

It is not easy to pinpoint the causes of these differences, but some contributory factors may be cited. There is, firstly, the question of the variation in definition of readership used. It is not impossible that while the perfunctory and casual reader of the newspapers would be qualified for inclusion in the IPA Survey, the Hulton question, requiring

TABLE II

Estimated proportion of male adult population reading certain newspapers and journals

	IPA Survey October to December 1953 Sample Size 7 880		Hulton Survey January to April 1954 Sample Size 6 116	Indices* of Readership	
	Percentage	Standard† Error × 1·5	Percentage	IPA	Hulton
Newspapers					
<i>Daily Mirror</i> ..	36	1·6	28	63	58
<i>Daily Express</i> ..	36	1·6	30	63	60
<i>Daily Herald</i> ..	22	1·5	12	38	24
<i>Daily Mail</i> ..	17	1·3	12	30	25
<i>News Chronicle</i> ..	12	1·2	9	22	18
<i>Daily Sketch</i> ..	7	1·0	4	13	7
<i>Daily Telegraph</i> ..	7	1·0	6	12	12
<i>The Times</i> ..	3	0·7	1	5	3
Sunday newspapers					
<i>News of the World</i> ..	58	1·6	49	100	100
<i>The People</i> ..	40	1·6	35	69	70
<i>Sunday Pictorial</i> ..	38	1·6	35	66	70
<i>Sunday Express</i> ..	23	1·5	21	40	44
<i>Sunday Dispatch</i> ..	18	1·3	16	32	32
<i>Empire News</i> ..	16	1·2	11	28	22
<i>Sunday Graphic</i> ..	8	1·0	6	15	12
<i>Sunday Chronicle</i> ..	6	0·7	5	11	11
<i>Reynolds News</i> ..	6	0·7	5	11	9
<i>The Observer</i> ..	4	0·7	4	6	9
<i>Sunday Times</i> ..	4	0·7	4	6	7
Weekly periodicals					
<i>Radio Times</i> ..	51	1·7	48	89	98
<i>Reveille</i> ..	28	1·5	27	49	54
<i>Picture Post</i> ..	13	1·2	16	22	33
<i>John Bull</i> ..	12	1·2	14	20	29
<i>Illustrated</i> ..	10	1·0	12	18	25
<i>Tit Bits</i> ..	9	1·0	11	15	23
<i>Week End Mail</i> ..	10	1·0	11	18	22
<i>Thomsons Weekly</i>					
<i>News</i> ..	6	0·7	7	11	15
<i>Everybody's</i> ..	7	0·7	9	13	19
<i>Punch</i> ..	2	0·5	4	4	8
Monthly periodicals					
<i>Reader's Digest</i> ..	9	1·0	15	16	32
<i>Lilliput</i> ..	3	0·7	4	6	8

* Expressed as a percentage of the readership of the *News of the World*, as estimated by the appropriate survey.

† Based on Tables of 'Standard error × 1·5' given in the IPA Survey.

that the informant should have *read* the paper and not merely *looked at* it, would have excluded him; the difference in the informant's mind between *looked at* and *read* could conceivably be less important in respect of weekly periodicals. But clearly one cannot be too certain to what extent questions specifying *reading* will exclude the casual reader more than questions specifying *looked at*. The Hulton Readership Survey of 1953, for example, points out that its definition of reading

has sometimes been criticised as being too wide, and as admitting 'casual' readership too freely as the equal of regular readership'.⁹ Furthermore, the discrepancy in the estimates relating to the *Reader's Digest* would not seem to be easily explained by the factor of definition. In considering the effect of the definition of readership on resulting estimates it might be found useful to analyse the data collected in the IPA Survey on the place where papers were reported to have been read. It would be interesting to see, for example, whether there are any significant differences in the proportion of publications read at home as compared with elsewhere, and whether these differences have any association with the scale of the variations between the Hulton and IPA estimates of readership of particular newspapers.

A second factor, and one which could seriously affect the results, is the use or absence of memory aids. In the IPA Survey aids were used in respect of each publication about which informants were asked. Informants were shown reproductions of the publication's heading when asked about it. Memory aids, in the form this time of 'check lists', were used in the Hulton Survey only when dealing with weekly and monthly publications. This, *prima facie*, is another possible reason for the closer agreement achieved between the two surveys with regard to weekly and monthly periodicals than has been noted above for newspapers, in respect of which the Hulton Survey does not apply 'prompt lists'. Research into the effect of memory aids, when used in different forms and to varying extents, on estimates derived from market research surveys of this type would clearly seem to be of considerable importance.

A further contributory factor which would tend to cause the Hulton estimates for weekly periodicals to be greater than those of the IPA Survey could have been the careful exclusion by the IPA technique of persons who would not have seen the paper 'a week ago yesterday', whereas those respondents might not always have been excluded from readership by Hulton. The effect of timing, with more weekly and particularly monthly periodicals being available in homes after Christmas (when the Hulton Survey was carried out) than before (when the IPA Survey was held), might similarly have caused Hulton estimates for such periodicals to be greater.

Yet another problem for investigation would seem to be the effect of the length of the questionnaire taken in relation to its design. In the IPA Survey respondents were asked a set of questions about each of some eighty publications. The intention was to ensure that only informants who had seen a copy of the publication in question during the qualifying readership period prescribed in the survey were included. This period was, for example, the previous day for daily newspapers, or the previous seven days for weekly publications. But these questions were asked in a fixed routine which could soon become known to the respondent. One wonders therefore whether some respondents, suffering from 'interview fatigue', might have realised that by answering negatively to the first question they would be excused from answering

the remainder of the set with respect to the publications concerned, and so could reduce the length of the interview. It may not be without significance that the differences between IPA and Hulton figures are greatest for newspapers, and are smallest for weekly and monthly publications which were 'asked about' last in both surveys.

The measure of this effect, if it existed, might usefully be extracted by rotating the order in which different publications were asked about, in addition to varying the order in which publications within any particular group were mentioned, as was done in all cases by the IPA and by the Hulton Survey in respect of weekly and monthly publications. In brief, when newspapers are asked about *first*, are resulting readership estimates higher than when they are asked about *after* the other types of publication? Or, if one asks about monthly periodicals *at the beginning* of the schedule, are the resulting estimates higher than if one asks about them *at the end*? Perhaps if these surveys are repeated this point could be examined, so that the effect of group order could be ascertained when either quota or random samples are employed.

Using Readership Surveys

How concerned the user of readership surveys is when confronted with the comparative results of the Hulton and IPA Surveys would, of course, depend upon the way in which the results are normally used. More information on this subject is badly wanted, and perhaps a market research practitioner will in due course describe in detail the mechanics of what is known in the advertising profession as 'media planning'—in other words, the way in which publication and other forms of advertising are chosen for use in particular advertising campaigns.

In view of the differences between the results of these surveys, it is possible that the planner may rely on firm data of numbers of copies of newspapers sold—in other words, the Audit Bureau of Circulation figures—but may use the readership surveys to subdivide the circulation figures according to the social class of purchasers. A comparison of the distribution of readership by social class and other characteristics is therefore relevant. The definitions of social classes used in these two surveys are not precisely comparable, but they are sufficiently similar not to invalidate the comparison for daily newspapers shown in Table III. In general the distributions are very close, although there are some marked differences. That one-third of *The Times* readership is, according to the IPA Survey, working class may sound a trifle odd; it could mean that 'public library' readership of such papers is high, but it may serve to emphasise the need to consider these estimates in conjunction with their sampling errors.

It is not unlikely, however, that in planning campaigns a comparison is made of the readership of one paper with that of another, in conjunction with the comparative costs of buying space in the respective newspapers. If this is so, the sort of problem confronting the user of

TABLE III

Distribution of male readership of daily newspapers by social class, as estimated by IPA and Hulton Surveys, 1954

				Well-to-do (%)	Upper and Lower Middle Classes (%)	Working Class and Poor (%)
<i>Daily Mirror</i>	..	..	IPA	1.4	19.8	78.8
			Hulton	0.8	15.0	84.2
<i>Daily Express</i>	..	..	IPA	3.5	26.2	70.3
			Hulton	3.8	29.1	67.1
<i>Daily Herald</i>	..	..	IPA	0.7	16.6	82.7
			Hulton	1.0	11.0	88.0
<i>Daily Mail</i>	..	..	IPA	6.3	30.8	62.9
			Hulton	7.0	38.4	54.6
<i>News Chronicle</i>	..	..	IPA	2.4	27.1	70.5
			Hulton	3.3	30.7	66.0
<i>Daily Sketch</i>	..	..	IPA	3.8	24.0	72.2
			Hulton	6.7	27.0	66.3
<i>Daily Telegraph</i>	..	..	IPA	18.4	51.4	30.2
			Hulton	23.7	63.9	12.4
<i>The Times</i>	..	..	IPA	28.2	40.0	31.8
			Hulton	NA	NA	NA

Definitions

'Well-to-do'—taken as 'Upper' in IPA Survey and as Class A in Hulton Survey.

'Upper and Lower Middle Classes'—taken as 'Middle' in IPA and as Classes B plus C in Hulton Survey.

'Working Class and Poor'—taken as these classes in IPA and as Classes D plus E in Hulton Survey.

'NA' signifies 'not analysed'.

This information may be demonstrated by comparison of the indices of readership included in Table II. Generally speaking, the differences in these indices, or *News of the World* parities as they might be called, are not very large, although they could possibly be greater if another publication were used as yardstick. One of the most marked differences relates to the *Daily Herald*. According to the IPA Survey, that paper's readership is 38 per cent of that of the *News of the World*; while according to Hulton its readership index is only 24 per cent.

In Table IV, readership figures by region of each of the national daily newspapers are again shown as indices in terms of the estimates for the *News of the World* as given in each survey. It will be noted that the IPA index is not always greater than the Hulton index. The biggest differences still relate to the *Daily Herald*, except in two northern regions. Further examination of the *Daily Herald* figures demonstrates that the difference between the IPA and Hulton Surveys applies to each of the social classifications used in these surveys. Table V shows the estimate of the proportions of adults reading the *Daily Herald*, according

TABLE IV

Regional indices of comparative readership by male adults
 ('News of the World' readership $\equiv$ 100 in each case)

	Total, All Regions	London and South- East	South- West and Wales	Mid- lands	North- West	North- East and North	Scotland
	IPA H	IPA H	IPA H	IPA H	IPA H	IPA H	IPA H
<i>Daily Mirror</i> ..	63 58	80 73	66 58	72 63	45 35	45 45	20 21
<i>Daily Express</i> ..	63 60	59 45	53 53	58 60	82 80	49 53	110 142
<i>Daily Herald</i> ..	38 24	32 22	46 24	43 25	39 36	51 26	16 16
<i>Daily Mail</i> ..	30 25	28 21	23 23	32 26	32 32	31 27	44 33
<i>News Chronicle</i> ..	22 18	27 22	17 17	20 12	28 23	19 17	6 3
<i>Daily Sketch</i> ..	13 7	16 7	13 6	14 11	12 10	7 4	6 NA
<i>Daily Telegraph</i> ..	12 12	23 19	13 9	7 9	6 8	5 6	2 NA
<i>The Times</i> ..	5 3	8 4	4 3	2 NA	4 3	2 NA	3 NA

'H' in the above Table is an abbreviation for 'Hulton'.

'NA' signifies 'not analysed'.

to these classes. It will be seen that in respect of each class the IPA figure is considerably greater than the corresponding Hulton figure

TABLE V

Estimated proportions of adult population reading the 'Daily Herald'

Social Class	Male		Female	
	IPA (%)	Hulton (%)	IPA (%)	Hulton (%)
Well-to-do (A)	4	2.5	3	1.5
Upper and lower middle classes (BC)	15	5.5	9	4
Working class (D)	25	15	17	10
Poor (E)	27		17	
All classes	22	12	15	8

For definitions of classes see notes to Table III.

A similar analysis made in terms of the age of respondents also shows that the IPA estimates are, especially in the case of the *Daily Herald*, consistently higher than the Hulton estimates; in respect of each of the age groups 16-24, 25-44, 45-64, and 65 and over, the IPA estimates of the proportion of respondents reading the *Daily Herald* tend to be much greater than the Hulton estimates.

In choosing the publications to be used in an advertising campaign the user accepts one set of comparative values or another; his choice will depend upon his own assessment of the results. The IPA Survey has the very real advantages of a random sampling design, and standard errors can be computed in respect of each estimate. However, the user will also need to consider problems unrelated to sampling, e.g. the effect of the definition of readership chosen, the way in which the type of interview might have influenced the results (as outlined above), and in particular the effect of memory aids, which might have been used in various ways or omitted entirely.

It is hoped that statisticians in the market research field will examine very carefully the differences in these two surveys in more detail than is possible from the published reports. Close study of the discrepancies within regions of readership estimates for different age and social groups may throw more light on the causes of some of the big differences in overall readership figures which emerge from broader comparisons.

Further Steps in Readership Research

The reviewer of these surveys must be impressed with the hard thought that went into their planning and execution. Indeed, the somewhat fragmentary comparative analysis made in the preceding paragraphs is intended only to suggest possible future lines of research on the basic measurement of readership. After this hurdle has been surmounted the specialist in this field has others to tackle: for example, there is the question of measuring the extent to which the insertion of advertisements in subsequent issues of a newspaper increases the potential audience for an advertisement. Then there is the problem of measurement of the effect of the position of an advertisement in a particular paper, and of course there are the thorny problems attached to the question of the effect of the design, layout and copy, or message, of an advertisement on its ability to attract attention and stimulate action. There is evidence that workers on market research are giving considerable attention to these and other problems.

The reviewer cannot fail to be struck by the quality of the presentation of most of the readership surveys. They also include data which are of interest to sociologists and business economists and which are often not available from official sources. On this point it is to be hoped that further work is being contemplated by the IPA on the mass of such data that was collected in the course of their enquiry; perhaps more replications in respect of sociological, economic, and demographic data will be included in subsequent reports.

But might I make the plea that when such results are published they should be incorporated in a less elaborate binding, so that the price at which the report is offered could be brought well below the ten guineas being asked for the IPA Readership Survey, and so bring it within the financial compass of statisticians and economists working in less remunerative fields than advertising?

Acknowledgements

I wish to acknowledge the benefit derived from discussions with a number of persons during the preparation of this paper. In particular, I am indebted to Mr C. A. Critchley, who read and commented in detail on an earlier draft. All views expressed and conclusions reached are, however, entirely my own responsibility.

REFERENCES

- ¹ SILVERMAN, RODNEY (1953). *Advertising Expenditure 1952*. The Advertising Association, London.
- ² ABRAMS, MARK (1954). 'The future of advertising', *Financial Times*, 15th December.
- ³ MARKET RESEARCH SOCIETY (1954). *Readership Surveys—A Comparative Study*.
- ⁴ WILKINS, L. T. (1952). 'Market research reports', *Operat. Res. Quart.*, **3**, 4.
- ⁵ ODHAMS PRESS LTD. (1953). *The Buying Habits and Purchasing Power of 'Daily Herald' Readers*.
- ⁶ KEMSLEY PRESS LTD. (1954). *A Survey of Certain Aspects of Newspaper Readership in the North of England*.
- ⁷ HULTON PRESS LTD. (1954). *The Hulton Readership Survey 1954*.
- ⁸ MOSER, C. A., and STUART, A. (1953). 'An experimental study of quota sampling', *J. R. Statist. Soc., A*, **116**, 349.
- ⁹ HULTON PRESS LTD. (1953). *The Hulton Readership Survey 1953*.

Reference not mentioned in the text

- HAYDOCK, J. (1954). 'Two readership researches—Why the differences?', *World Press News*, 6th August 1954.

THE DETERMINATION OF BRANCH STAFF ESTABLISHMENTS IN A MULTIPLE SHOP ORGANISATION

BERNARD T. RAMM

Mr Ramm shows how elementary queue theory can be applied to the results of time study in order to determine the number of staff required to give a satisfactory standard of service in a shop.

THERE are two main ways in which the head office of a multiple shop organisation may control the numbers of staff at the branches. The first, and more usual, method is to allow the area superintendents, or even the branch managers, to employ the staff they consider necessary. Control by head office is then exercised by means of an examination of the percentage of takings spent on wages at each branch. This method has the disadvantage of bearing heavily on the small branch, which in general cannot be run as economically as the larger one, and reducing the standard of service there. A further disadvantage in the pre-war days, when there was a buyers' market in labour, was that it often resulted in the more highly paid long-service employee being discharged in preference to the cheaper youngster.

The alternative method of control is to draw up establishment scales so that the size of a branch's trade automatically determines the number of staff required to provide a uniform standard of service, the standard being decided according to what the company feels it should give the public in conjunction with its cost as a whole rather than at individual shops. The subject of these notes is the drawing up of such scales for the staff of a firm of multiple grocers by means of the application of elementary queue theory to observations derived from time studies.

Originally these scales, a separate one for each counter in the shop, were rough and ready ones drawn up after discussion round a table between the area superintendents. In the rapidly changing conditions of the post-war years they needed revision, and the opportunity was taken to get the superintendents, instead of talking in abstract terms, to give independently their views on the necessary staff for a number of specified branches of varying size. The estimated requirements for each counter were then plotted against weekly sales at the specified shops and a straight line was fitted to the points, from which a scale of establishments was drawn up. It was immediately apparent that each superintendent had his own ideas on what constituted a satisfactory staff, and that they held different opinions on the way in which it could vary with the amount of business done. Furthermore it was

impossible to say what was the standard of service given. It was therefore decided to draw up new scales based scientifically on time studies of the actual serving of customers.

To do this it is necessary to know much more than the mean service time and the number of transactions per week. Customers do not arrive in equal numbers on each day of the week, but their shopping habits follow a well-defined pattern which is fairly constant from week to week. Superimposed on this are random fluctuations from minute to minute in their arrival rate, and similar variations in the length of time it takes to serve them. Staff time is therefore wasted both by the uneven arrival of customers taken hour by hour over the week and by random variations in their arrival and service time. The allowance to be made for the first of these factors can be determined by observations of the pattern over the week of customer arrivals; for the second, recourse has to be had to the mathematical theory of queues.

To determine the amount of waste time and calculate an adequate number for the staff the following information is needed: (a) the mean arrival rate of customers at hourly (or even half-hourly) intervals throughout the week in relation to the total weight or value of goods sold over the counter each week; (b) details of the mean and frequency distribution of the service time for each customer; and (c) information on the behind-the-scenes preparation work required. All these details can be found by observation in a representative sample of branches. In the case of (b) and (c) the observed times for each of a number of operators working under busy conditions were rated up or down to a standard efficiency which may be expected from trained staff. In the case of (a) the observations were confined to the busy part of the week and expressed as an arrival rate per hour divided by the weight of goods sold per week in tons on counters where the bulk of the sales consist of a single item, or divided by the value sold per week in £100's where they comprise numerous different items. It must then be decided whether the shops should be staffed to meet the arrival rate, say, in the peak half-hour of the week or only perhaps that of the peak two hours. Let this rate be m customers per hour per ton sold each week.

A decision must also be taken as to the standard of service to be offered. This is most conveniently expressed as an average waiting time measured as a multiple of the average service time. Let the mean waiting time in minutes be $\bar{t}_w$ and the mean service time in minutes $\bar{t}_s$; then $\bar{t}_w = k\bar{t}_s$, where k is a constant determined by the standard of service to be offered during peak periods. If the proportion of wasted time in these periods due to random variations in the time interval between customers is $(1 - r)$, then each server can, on average, handle $60r/\bar{t}_s$ customers per hour, equivalent to a weekly sale of $60r/m\bar{t}_s$ tons (or £100's) of goods. It remains to determine r .

It is clear that a large number of servers can, by co-operating with one another, iron out variations in the rate of customer arrivals, and r will tend to unity as the number of assistants behind a single counter

increases. For a single server it is not difficult to make an estimate of r , using the relations developed by D. G. Kendall.¹ Provided that equilibrium has been reached—and the arrival rate must not be so large that this is impossible—then for a single server r is given by the formula

$$r = \frac{2k}{1 + \frac{\sigma^2}{\bar{t}_s^2} + 2k}$$

where σ is the standard deviation, determined by actual time studies, of the service time. In practice it was found that if the standard of service provided was such that the average waiting time was equal to the average service time, i.e. $k = 1$, then for a single server r had values for different counters varying from 0.6 to 0.8.

We have, then, two limiting values for r : the upper limit of unity and lower limit calculated for a single server. In view of the fact that the numbers of staff and size of trade can never be matched with very great accuracy (e.g. the number of staff must always be integral; and variations occur in the efficiency of individual workers as well as differences between branches in the intensity of shopping peaks) it was considered that only a rough estimate of the way in which r varied between its limits for more than one server was sufficient. These estimates were checked (and some adjustments made) by a small number of sets of observations of the operation of counters with several servers. These entailed noting the arrival rate of customers and recording at five-minute intervals the number of customers waiting to be served and the number of servers.

Once these values for r have been determined it is a simple matter to draw up a scale giving the weekly sales handled by one, two, three, or more assistants. This scale will only give the number of staff required for the actual sale of the goods to the customer. In addition to the time spent serving customers allowance must be made for the time spent in preparatory work. Some small operations can be done in slack periods, even on a busy day, and no allowance need be made for these; others, such as the boning and rashing of bacon, are major operations for which full allowance has to be made. Since, however, several hours' stock of goods prepared for sale may be carried, the problem is much simpler than that for counter staff. No allowance need be made for random minute-to-minute fluctuations in the arrival rate of customers but solely for the diurnal variation in sales; a figure is then worked out for preparation work (in the form of x tons per week per head) based on the staff needed during the busiest day of the week—usually Friday. If n operators can handle a total of T tons a week, considering only service and not preparation, then the overall staff requirement for this quantity of goods is $n + T/x$. The next step is to bring the numbers of staff in this scale back to integers by interpolation, so that finally we finish with a scale of establishments for each counter showing the range

of sales that can be handled by one, two, three or more staff. The increment in the quantity handled for each extra assistant will be an increasing function of the number of staff as the proportion of waste time decreases and the value of r approaches unity.

After establishment scales have been calculated in this way it remains to test their cost against the amount of money available. This may call for some adjustment to the standard of service laid down and the calculation of fresh scales. There is an important difference between this comparison and controlling on cost alone, in that the exact effect on the service offered to the customer can be calculated and false economies avoided.

The scales cannot always be rigorously applied, because the resulting numbers of staff will in some cases need adjustment to suit local conditions: e.g. especially intense trading peaks, premises too small to take the requisite number of staff, or so large that extra staff are needed.

It must be emphasised that the scales drawn up in this way rest entirely on the assumed peak rate of arrival of customers in relation to the shop's weekly trade. This was based on observations taken under actual staffing conditions, which must influence the arrival rates; in a badly staffed shop it is the peak trade that is the first to suffer. This difficulty was met as far as possible by avoiding branches known to be short-staffed. In addition, some discretion was exercised in choosing the observations included in the mean rate actually used. Without these precautions the final result may be unduly influenced by the establishments already in existence.

REFERENCE

- ¹ KENDALL, D. G. (1951). 'Some problems in the theory of queues', *J. R. Statist. Soc.*, B, **13**, 151.

A DESIGN FOR TESTING SEVERAL TREATMENTS UNDER CONTROLLED ENVIRONMENTAL CONDITIONS

GEOFFREY M. CLARKE

University of Bristol, Long Ashton Research Station

Many experimental tests are carried out in controlled environments, but variations within the environment are usually found, even with the most careful control. In this article Mr Clarke shows how a design based on Graeco-Latin squares was used to balance out the most important trends in a laboratory incubator.

MANY writers have drawn attention to the limitations of the Graeco-Latin square design, which is able to remove four orthogonal sources of variation in an experiment, but in doing so reduces considerably the degrees of freedom (d.f.) available for the estimation of error. Thus in a single 4×4 square, after taking out terms in the analysis of variance for rows, columns, Latin letters, and Greek letters, each with 3 d.f., we have only 3 d.f. for estimating the error, and this number is totally inadequate. This article indicates how four such squares can be used together, provided certain conditions of orthogonality are satisfied, in a practicable and precise experimental design.

Fungicide Assays in Incubators

In a series of bio-assays based on work reported elsewhere,² fungal mycelium, with fungicide incorporated in the nutrient medium, is incubated in Petri dishes, positioned in a standard incubator according to the layout described below. Growth is measured after two and four days, and the actual figure analysed is the growth between the second and fourth day. In each experiment, sixteen treatments are compared, each being replicated four times. One of these treatments will usually be a standard, and different fungicides can then be related to one another via this standard, which is subject to variation from run to run. This variation appears to be an unavoidable characteristic of assay work of this type.

Factors which may affect the fungal growth, apart from the fungicides themselves, are

- (i) vertical differences in the incubator,
 - (ii) horizontal trends through the incubator,
 - (iii) order of pouring out of fungicide.
- (iv) When dishes are stacked in vertical layers in an incubator, it is often found that there is a differential degree of growth of their contents at

various levels which is quite independent of treatments applied. Most frequently the dishes in the upper layer show significantly better growth rates than all others. This phenomenon has been observed in some few of the present trials; it is not possible to attribute it to any obvious cause, though it may be connected with slight air currents in the incubator. It is also conceivable that volatile fungicides escape somewhat more rapidly from dishes on the top layer. In this series of trials, dishes were stacked on a shelf half-way up the incubator, reaching from there almost to the top. If a layer of dishes rests on the floor of the incubator, differential growth between the lowest and other layers may also be expected.

(ii) Owing perhaps to slight imperfections in the incubator, there may exist trends either from front to back or from side to side. There was no reason to suspect such effects in the incubator used; however, as a measure of safety, alternate sets of ten runs were laid out so as to examine trends from front to rear, the other sets of ten examining side-to-side effects. In the latter runs, growth has frequently proved significantly higher on the left side; this again might have been due to air currents, since a slight warp in the inner door of the incubator on the left side was discovered on closer inspection. No effect has been observed in a direction from front to rear.

(iii) The fungicide is made up as an emulsion, the proper amount of which is pipetted into the nutrient medium, and the whole agitated; there is a possibility with the type of fluid used that the four samples poured from this nutrient medium may not be perfectly uniform.

Typical Design

A typical design ready for use is as follows:

Layer IV	C ₁	J ₃	E ₄	R ₄	P ₁	B ₂	K ₁	F ₂
III	N ₂	H ₁	B ₄	M ₁	Q ₃	L ₃	G ₃	C ₃
II	Q ₁	D ₂	K ₂	F ₃	E ₁	J ₄	A ₁	N ₄
I	L ₄	P ₃	G ₂	A ₃	M ₂	D ₄	H ₄	R ₂
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Square	W				X			
Layer IV	N ₃	H ₃	A ₄	L ₂	D ₃	Q ₂	M ₄	G ₁
III	D ₁	K ₄	R ₁	E ₂	J ₂	P ₄	F ₄	A ₂
II	P ₂	M ₃	B ₃	G ₄	C ₄	H ₂	L ₁	R ₃
I	Q ₄	F ₁	J ₁	C ₂	B ₁	E ₃	K ₃	N ₁
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Square	Y				Z			

layer IV is the uppermost in the incubator, and 'squares' W-Z are placed according to the plan

Front	(4)	(8)	(12)	(16)	Back
(or Left)	(3)	(7)	(11)	(15)	(or Right)
	(2)	(6)	(10)	(14)	
	(1)	(5)	(9)	(13)	
	W	X	Y	Z	

The fungicides are allotted letters A to R at random (I and O being omitted to avoid confusion with numerals), the figures 1 to 4 indicating order of pouring of fungicide. It can easily be seen that each fungicide occurs once in each square and once in each layer, and that each pouring order is represented four times in each square and each layer. Thus the effects 'squares', 'layers', 'pouring', 'treatments' (i.e. fungicides) are mutually orthogonal⁵ (p. 49). In this design, orthogonality of horizontal trend to treatments and to order of pouring is possible in one direction only, so that 'squares' can be arranged to take out a horizontal trend *either* from front to back *or* from side to side of the incubator. In (ii) above it is indicated how the design has been applied in an experimental investigation^{2,3}; if a real trend does exist in one direction, 'squares' must of course always be arranged to eliminate this. A further condition for the use of this design is that there must be no square-layer interaction. These restrictions do not prevent the design from controlling all necessary sources of variation in the work for which it was devised; there is no evidence that parts of a single layer will behave differently from other parts so as to give rise to any square-layer interaction, any differences observed in growth from layer to layer being characteristics of complete layers.

It is possible to eliminate horizontal trends in both directions at the expense of allowing the treatments to fall into four groups of four, so that each group of four letters always occurs together in the same vertical stack of dishes. Although such a design, satisfying all the requirements of orthogonality, may seem preferable in theory, and appears to have an advantage if considerable horizontal variation is expected, it was felt at the planning stage of the work³ that any tendency to systematic grouping of the treatments should be avoided, especially since treatments frequently form factorial sets, as in the example described below. Further, though experience suggested a real vertical effect, the horizontal trends were examined as a subsidiary, interesting feature; the trend which has appeared is most likely a defect in the structure of the particular incubator rather than a general rule.

The alternative hypothesis is made that the growth y_{ijkl} is composed of a mean growth m , a treatment effect t_i , a layer effect l_j , a 'squares' effect s_k , and a pouring effect p_l , according to the additive model

$$y_{ijkl} = m + t_i + l_j + s_k + p_l + e_{ijkl}$$

where the e 's are all independent and are distributed normally with zero mean and variance σ^2 . The null hypothesis is

$$y_{ijkl} = m + e_{ijkl}$$

and the following easy analysis of variance is possible, sums of squares being calculated in the usual manner for orthogonal designs:

Item	D.F.
Squares	3
Layers	3
Pouring	3
Treatments	15
Error	39
Total	63

A simple analysis of this type is a valuable attribute of a design for routine work. Significance tests can be carried out, 'experimental error' estimated, and treatment means compared, by methods which are easy extensions of those for Latin and Graeco-Latin squares¹ (Chapter 4). In the work mentioned previously,^{2,3} the treatment means had to be estimated and compared in appropriate groups corresponding to their chemical structure; in addition significance tests were required to examine incubator trends. The procedure is illustrated more fully below.

Construction of Design

The basis of the construction of the design is to use four 4×4 Graeco-Latin squares 'tied', that is, laid side by side to form a 4 row $\times$ 16 column array⁶ (§ 24), the rows and columns subsequently being randomised. There are only two standard 4×4 Graeco-Latin squares,⁴ namely

A α	B β	C γ	D δ	and	A α	B β	C γ	D δ
B γ	A δ	D α	C β		B δ	A γ	D β	C α
C δ	D γ	A β	B α		C β	D α	A δ	B γ
D β	C α	B δ	A γ		D γ	C δ	B α	A β

Four squares are written down, the choice from these two standards being made at random, to give a basic layout of this form:

A α	B β	C γ	D δ	E α	F β	G γ	H δ	J α	K β	L γ	M δ	N α	P β	Q γ	R δ
B γ	A δ	D α	C β	F δ	E γ	H β	G α	K δ	J γ	M β	L α	P γ	N δ	R α	Q β
C δ	D γ	A β	B α	G β	H α	E δ	F γ	L β	M α	J δ	K γ	Q δ	R γ	N β	P α
D β	C α	B δ	A γ	H γ	G δ	F α	E β	M γ	L δ	K α	J β	R β	Q α	P δ	N γ

1 2 3 4

w

5 6 7 8

x

9 10 11 12

y

13 14 15 16

z

example of such an analysis follows, this particular experiment having been reported in detail³ under the code title 'material C'. Each individual result is based on the average of four diametric measurements of the growing area. The growth totals, expressed to the nearest 0.5 mm, are given in Table I.

The analysis of variance is:

Item	D.F.	S.S.	M.S.	
Squares ..	3	2.480 5	0.826 8	$F_{(3, 39)} = 3.07^*$
Layers ..	3	2.949 2	0.983 1	$F_{(3, 39)} = 3.65^*$
Pouring ..	3	1.261 7	0.420 6	$F_{(3, 39)} = 1.56$ n.s.
Treatments	15	295.339 8		
Error ..	39	10.496 1	0.269 1 ($= \sigma^2$)	
Total ..	63	312.527 3		

* Significant at 5 per cent level.

n.s. = not significant.

The F -tests enable any incubator trends or order-of-pouring effects to be examined; and it is seen that layer IV shows a higher growth than I, II and III, sufficient to cause significance of $F_{(3, 39)}$, that a low growth value in square X is likewise noticeable, and that pouring orders do not differ significantly. The layer effect is as suspected, and has been reproduced frequently in similar routine incubator work; square X here displays an effect not reproduced on any other occasion (in some thirty routine experiments), though a 'squares' effect frequently does show up as a high value in square W when the squares are laid from left to right of the incubator. Hence when using this particular incubator, squares are now always placed in this way.

There is no need of an F -test for treatments; in any case the mean square clearly is a large one. The treatments consisted³ of five substances α , β , γ , δ , ϵ , each at three concentrations (200, 100 and 50 parts per million), and a control (untreated). (This use of Greek letters has no connection with their earlier use in the squares.) These treatments corresponded to the following letters in the randomisation: $\alpha_{200} = K$, $\alpha_{100} = M$, $\alpha_{50} = J$; similarly $\beta = B, R, C$; $\gamma = G, H, E$; $\delta = L, D, A$; $\epsilon = Q, N, F$; and control = P. It was required to compare α_{200} with α_{100} with α_{50} , and so for β , γ , δ , ϵ ; α_{200} with β_{200} with γ_{200} with δ_{200} with ϵ_{200} , and so for 100, 50; and to observe which treatments did not differ significantly from control, i.e. proved completely ineffective as inhibitors. In order to make these comparisons most easily, the 'significant differences' between means are calculated; these are the numerical sizes of differences which must be exceeded to show significance at the usual levels (5, 1, and 0.1 per cent) and are given by the formula $t\sqrt{(2\sigma^2/r)}$, σ^2 being the error mean square, $r (= 4)$ the number of replicates of each treatment, and Student's t having 39 d.f. and the appropriate probability level ($P = 0.05, 0.01,$

001). Significant differences for this example are 0.742 (5 per cent), 0.994 (1 per cent), 1.307 (0.1 per cent), and these lead to the conclusions that with all substances except ϵ , increasing concentration significantly increases inhibition, that $\epsilon 100$ and $\epsilon 200$ cannot really be distinguished, and that $\epsilon 50$ and $\beta 50$ are not conclusively better than the control. (It should be noted that the corresponding significant differences for the treatment totals in Table I are 2.968, 3.976, and 5.228; i.e. roughly 2, 4, and 5.) A comparison of substances splits them up into two very well-defined groups, namely α , γ , δ and β , ϵ . Comparable concentrations of α , γ , and δ do not differ significantly from one another; and although at the high concentration β appears a better inhibitor than ϵ , this is not upheld for 100 and 50 p.p.m. Both β and ϵ are very much less effective inhibitors than α , γ , and δ .

Since the control growth (treatment P in this example) may vary from run to run of the series of experiments, it is convenient for tabulation to work in terms of 'percentage inhibition' when describing a volume of work of this type³; for a particular run, each treatment growth is related to control ($\equiv 100$ per cent), and the result subtracted from 100 to give a 'percentage inhibition'. This process aids in examining repetitions of trials, or different sets of concentrations included in different trials, or in comparing in an approximate way compounds which have not been included in the same run; each particular run is first examined in detail as described in the previous paragraph, and must be designed to give answers to specific problems as above, rather than to include an unplanned selection of available compounds.

Extension to Fruit Storage Experiments

It is well known that different varieties of apples, etc., or one variety subjected to different treatments in the orchard, have different storage characteristics, keeping for varying lengths of time and ultimately displaying different types of breakdown.⁷ In order to compare varieties, or different orchard treatments, samples of fruit are laid out in store and examined at various times during the winter. Factors for which allowance may be needed in the subsequent analysis are:

- (i) vertical differences in the store,
- (ii) horizontal trends through the store,
- (iii) orchard effects.

(i), (ii) Vertical variation is very probable in a large fruit store, in the same way as it has been observed in the more easily controlled environment of an incubator, and a horizontal component of variation may likewise be expected. These will be due to slight differences in temperature and humidity, and the effects of artificial ventilation.

(iii) Samples will frequently be taken on a block-to-block basis corresponding to the layout of the experiment in the field from which the fruit is taken. This block effect will be treated in the same way as the 'pouring' effect above. It will sometimes occur that samples come

from different sources, not merely from blocks, and this can similarly be allowed for. Further, the sixteen 'treatments' above could be partitioned to form, for example, an 8×2 set of eight orchard treatments at two sites, four blocks (equivalent to 'order of pouring') being used from each site. The eight orchard treatments might be, for example, a 2^3 factorial NPK set. (Sites, being a non-randomised factor, could not be tested for significance.)

Other Extensions

The above design is based on 4×4 squares; similar useful layouts can be obtained using 3×3 squares for 9 treatments, but larger layouts, e.g. using 5×5 squares, are likely to be too unwieldy to be of use in laboratory incubator work, though this amount of material may well be involved in other extensions. Where the 4×4 -square design is used with four treatments only, it of course reduces to the simple 'tying' of squares.⁶

Acknowledgements

I wish to thank Mr G. V. Coles and Dr R. J. W. Byrde for their collaboration in the use of this design in incubator work, and Mr R. W. Marsh for a suggestion concerning other applications.

REFERENCES

- ¹ COCHRAN, W. G., and COX, G. M. (1950). *Experimental Designs*. Wiley, New York.
- ² COLES, G. V., and BYRDE, R. J. W. (1954). 'The fungicidal properties of coal tar distillates: Progress report', *Ann. Rep. Long Ashton Res. Sta.* 1953. 109.
- ³ COLES, G. V., MARTIN, J. T., and BYRDE, R. J. W. (1955). 'The fungicidal properties of coal tar distillates: Progress report II', *Ann. Rep. Long Ashton Res. Sta.* 1954. 74.
- ⁴ FINNEY, D. J. (1945). 'Some orthogonal properties of the 4×4 and 6×6 Latin squares', *Ann. Eugen., Lond.*, **12**, 213.
- ⁵ KEMPTHORNE, O. (1952). *The Design and Analysis of Experiments*. Wiley, New York.
- ⁶ PEARCE, S. C. (1953). *Field Experimentation with Fruit Trees and other Perennial Plants*. Tech. Comm. No. 23, Commonwealth Bureau of Horticulture, East Malling, Kent.
- ⁷ WALLACE, T. (1953). *Science and Fruit*, p. 140. University of Bristol.
- ⁸ YATES, F. (1953). 'The analysis of replicated experiments when the field results are incomplete', *Empire Journal of Experimental Agriculture*, **1**, 129.

NOTES AND COMMENTS

It is with great sorrow that we record the death, on 15th July 1955, of Mr Philip Lyle. He was a director of Messrs Tate and Lyle Ltd and did work of great influence in the field of sugar technology. Statistics, although perhaps a subsidiary interest, was a lively one, and on applied statistics he has left his mark. His well-known book, *Regression Analysis of Production Costs and Factory Operations*, is more than a work of exposition; and his articles on 'The Construction of Nomograms for Use in Statistics' in the third volume of *Applied Statistics*, with their emphasis on 'empirical nomograms', present an unusual and fruitful approach to the subject. Philip Lyle was an Honorary Secretary of the Royal Statistical Society and an active member of Council and several committees; he was a keen supporter of the work of the Sections of the Society and was very appreciative of the life and vigour given to statistics by the work of the younger people. Behind the scenes he was at the centre of the moves that led to the foundation of *Applied Statistics*, and his advice and support contributed much to its successful start. Those who worked with him were warmed by his kindly and modest charm and by his willingness to help in any way possible. Besides a wife, a daughter, and two sons, Philip Lyle has left many friends to mourn his loss.

★ ★ ★

During the past year Professor Paul C. Clifford of the USA has been in Europe acting as a consultant to the European Productivity Agency (EPA) on quality control. He has visited many Western European countries, stimulating the development of quality control and of training in the subject. In some countries he found the subject to be well developed and widely applied; in others very little had been done. As a culmination to Professor Clifford's term of office a conference was held in Paris on 11th-13th July 1955 at which there were about a hundred delegates from thirteen countries. There were papers on the organisation of quality control in the factory, on training, on applications to particular fields, and on statistical techniques. Statistical topics predominated, but the subject was throughout treated as a practical, industrial subject. The papers and discussions were lively and interesting, but perhaps the chief outcome of the conference was something like the birth of a corporate quality control movement in Western Europe. It is likely that the EPA will take steps to foster this movement.

★ ★ ★

The Study Circle on Medical Statistics, which has hitherto been included in the Study Section of the Royal Statistical Society, was in July 1955 accorded the status of a Section of the Society. It is now known as the Medical Section. We are pleased to note this recognition of the important part played by statistics in the medical field.

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

List and Reports of Meetings

BIRMINGHAM AND DISTRICT GROUP

2nd March 1955. Some Local Statistics: R. Padley

Dr Padley, who is Statistical Officer to the City of Birmingham Corporation, showed how his department serves the various local government committees. He quoted examples of statistical investigations and demonstrated the results by maps.

6th April 1955. What an Industrial Applications Section should talk about
An informal discussion on this subject was opened by Mr B. J. A. Martin.

GLASGOW GROUP

10th March 1955. Forecasting Sales: Miss J. Keen

LEICESTER GROUP

24th March 1955. A Practical Approach to making Statistical Decisions:
G. A. Barnard

See the report in *Applied Statistics*, 4, 124.

LONDON GROUP

4th March 1955. An Application of Monte Carlo Methods to Communication Problems: J. C. R. Clapham

Some telephones in collieries may possibly fail to answer an emergency call, although other telephones may do so. In such situations it is not straightforward to calculate how long it will take to warn men in danger, since the route of the warning message is uncertain. Mr Clapham showed how the problem could be tackled by a Monte Carlo method.

1st April 1955. Some Recent Advances in Experimentation to attain Process Improvement: G. E. P. Box

Most chemical systems have been found to give yield contours corresponding to a stationary or rising ridge rather than to a simple peak. This is probably because the contours are not plotted for truly independent variables such as the number of molecular collisions per second, but rather for dependent variables such as temperature and concentration. A second difficulty in this work is the interpretation of complex second-order polynomials. However, conversion to the canonical form often leads to useful information. Dr Box maintained that process improvement work should pass beyond the stage of significance testing and should have as its goal the understanding of the basic reaction mechanism.

15th May 1955. Statistical Methods in Metallurgy: U. R. Evans

Dr Evans described the use of statistical methods to assess the risk of some unlikely, and usually undesired, event such as fracture, perforation, or failure to fulfil specification. He quoted cases where histograms were misleading and described the use of probability paper, which, despite some disadvantages, has many useful applications in metallurgy.

1st June 1955. Visit to Kodak Works and Laboratories, Harrow

A tour of the factory was followed by tea, after which Mr A. K. Soper gave a talk on the application of statistical methods to problems of photographic processing.

MERSEYSIDE GROUP

15th March 1955. Linear Programming: N. Williams

Mr Williams discussed the methods of solving simultaneous equations and showed how the Simplex method worked in simple cases. The solution obtained was illustrated graphically. Among the practical examples discussed was Koopman's transportation problem.

19th April 1955. Particle Size Distributions: G. P. Davies

In order to burn a fuel efficiently the surface area which comes into contact with oxygen must be a maximum; this is achieved by various methods of atomisation. Their efficiency is assessed by estimating a parameter known as the Sauter Mean Diameter (SMD), which measures the ratio of the volume to the surface area of the droplets comprising the spray. Mr Davies described investigations into techniques for estimating the SMD. Various frequency distributions were fitted to the results and it was found that the customary χ^2 test of goodness of fit was quite unsuitable for assessing which distribution yielded the best estimate of the SMD. Several distributions were therefore fitted to a number of counts where the SMD had been carefully determined by a method which is unsuitable for routine work. The SMD's estimated from these distributions were compared with the accurately determined values. A distribution proposed by Mugele and Evans (*Industr. Engng Chem.*, **43**, 1317) was found to be the most suitable.

NORTH-EASTERN GROUP

1st March 1955. Probits and Logits: R. L. Plackett

SHEFFIELD GROUP

1st March 1955. The Bunching of Furnaces: R. Solt

The variability of furnace cycle times leads to occasions when a number of furnaces become ready to be charged at the same time. When such 'bunching' is severe it may be impossible to charge the furnaces at the required rate. Mr Solt showed charts giving the incidence of charging and the resulting shop charging-rate for a five-furnace melting shop. The intervals between successive furnaces beginning to charge were recorded and a frequency distribution was developed to show the probability with which a given proportion of furnaces would begin to charge at the same time. If the charging system is to meet the demands of the shop, the storage capacity and replacement capacity of the system must be properly balanced. The

normal charging time and the average charging rate for the furnaces were considered in relation to the various data concerning the supply system. From this a 'critical time' was developed, defined as the shortest interval between the first and last furnaces beginning to charge that will allow charging to proceed without delay. The bunching distribution was related to this critical time and an average delay per charge was developed for any given set of circumstances.

28th April 1955. Labour Turnover Analysis: J. E. Andrew

In the method of analysis used the probability of an employee leaving was expressed as a function of his length of service. Dr Andrew gave two methods of estimating this function: one involves the following-through of a sample of entrants, giving a survival curve, whereas the other uses census data in arriving at a stability curve. In practice both curves were found to be approximately lognormal, and this assumption simplified the calculation of an expectation of service, which was proposed as a single measure of labour stability. Examples were given of differences in stability between institutions, and between departments within an establishment, and the effect of living circumstances and age upon the probability of leaving was described. Finally the tests of goodness of fit to the lognormal distribution were dealt with.

SOUTH WALES GROUP

18th March 1955. Estimating the Number of Children of Employees at a Works: J. F. Cutts. A Problem of Weighing: W. W. Jones. The Approximate Analysis of some Non-orthogonal Experiments: W. D. Ewan

Mr Cutts described a method used for estimating the number of children of employees, based on the distribution of income-tax codings. Assumptions about the maximum expected positive and negative skewness in the different humps of the distribution enabled upper and lower limits to be calculated.

Mr Jones discussed a weighing problem, in connection with the sale of chemicals by standard weight, where the container also varied in weight appreciably.

Mr Ewan described an approximate method of analysis involving little computation which he used in weighing-trials when the experimental design ended up as unbalanced paired comparisons. The approximate analysis, including the calculation of the errors of treatment comparisons, was found to be remarkably close to the exact analysis in all cases checked.

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

List and Reports of Meetings

BRISTOL GROUP

17th March 1955. Retail Trade Statistics: J. Stafford

Mr Stafford outlined the type of information available and its sources, explained the use made of this information by his department at the Board of Trade, and described the figures produced from these statistics.

STUDY CIRCLE ON MEDICAL STATISTICS *

12th January 1955. Statistics in Parapsychology: R. H. Thouless
Joint meeting with the Study Section.

22nd February 1955. Medical Field Surveys: A. L. Cochrane

Dr Cochrane described some findings of a follow-up survey in the Rhondda Fach which was carried out in 1953 by the MRC Pneumoconiosis Research Unit. He first discussed the value of such indices as mortality, attack rate, and prevalence rate, and pointed out that, in a period of rapidly falling mortality, the prevalence rate might well increase. With his colleagues he had X-rayed 91 per cent of the population of the valley over the age of 5 (some 25 000 people) and by comparison with a similar survey carried out two years previously he was able to demonstrate a marked fall in the prevalence of infectious tuberculosis among women and younger men, but a slight increase in older miners and ex-miners.

29th March 1955. Hospital Statistics: C. J. Donelan, K. W. Cross, and B. Abel Smith

The three speakers discussed this problem from the respective viewpoints of planners in the Ministry of Health, a Regional Hospital Board, and an economic sociologist.

26th April 1955. Some Hypotheses on Coronary Artery Disease: J. N. Morris

Various hypotheses have been advanced to account for the recent increase, which appears to be largely confined to the Anglo-Saxon countries, in mortality ascribed to coronary artery disease. Dr Morris summarised the evidence for and against some of these hypotheses and outlined some of the work he had carried out with the MRC Social Medicine Research Unit into the occupational aspects of the problem.

31st May 1955. Controlled Trials in Clinical Medicine. Sequential Methods: P. Armitage. Some Practical Problems: J. Knowelden. Statistical Inference: R. F. Wrighton

*Now the Medical Section.

BOOK REVIEW

Cedric A. B. Smith: *Biomathematics: The Principles of Mathematics for Students of Biological Science*. London: Charles Griffin, 1954. Pp. xv + 712. 80s.

This book is the third edition and a rewritten version of the work by the late W. M. Feldman. It contains a full exposition of the statistical methods used in biology. It is said to be equally suitable as a revision course for post-graduate workers and as a textbook for students, since it leads up from elementary mathematics to all the advanced techniques likely to be required by the research worker. It is not easy for the mathematician to realise that what to him is childishly elementary may be hopelessly complicated to the biologist who is not mathematically inclined. If the book is intended as a working handbook for students of this type it does not fulfil its purpose. The text is said to be 'enlivened' by imaginary dialogues. These the reviewer finds coyishly heavy and unhelpful. They might well be omitted in the next edition. The book contains many useful statistical techniques not too hard to apply and should come into wide use as a standard work of reference.

S. G. HARLAND

PUBLICATIONS RECEIVED

Martin Ekblad: *Induced Abortion on Psychiatric Grounds—A Follow-up Study of 479 Women*. Translated into English by Donald Burton. Copenhagen: Ejnar Munksgaard, Nørregade 6, 1955. Pp. 238.

Económica—Revista de la Facultad de Ciencias Económicas. Argentine: Universidad Nacional de Eva Peron, published three times yearly. In Spanish, with summaries in German, French, English and Italian. \$20 m/n.

T. Ghose: *Survey of Retail Markets in Certain Selected Areas of West Bengal*. Calcutta: Bureau of Industrial Statistics, 1955. Pp. 118.

Harold C. Edey and Alan T. Peacock: *National Income and Social Accounting*. London: Hutchinson's University Library, 1954. Pp. 224. 8s. 6d. (Referred to in the article commencing on p. 133.)

Institute of Practitioners in Advertising: *The National Readership Survey—1954 Edition*. London, 1954. Pp. 115, 45 tables. £10 10s. (Reviewed in the article commencing on p. 183.)

CORRIGENDUM

The Design of Social Research, by R. L. Ackoff, is published by the University of Chicago Press, and not by the Cambridge University Press as stated in our June issue. The Cambridge University Press are the agents in this country.

APPLIED
STATISTICS

INDEX

- accountant, and statistics, 68
- accumulation, calculation by, 90
- Ackoff, R. L.*, 128, 212
- analysis of dispersion, 113
- Andrew, J. E.*, 210
- Association of Incorporated Statisticians, 121
- Barnard, G. A.*, 124
- Bartlett, M. S.*, 131
- Bartlett's decomposition theorem, 106
- Baer, S.*, 1
- Bennett, C. A.*, 129
- Bernstein, L.*, 70
- Billewicz, W. Z.*, 97
- 'Biomathematics', review, 212
- 'Biometrika Tables for Statisticians', review, 72
- Birmingham, 'Abstract of Statistics', notice, 132
- birth weights, 97
- Blackett, P. M. S.*, 67
- Boswell, D.*, 123
- Box, G. E. P.*, 125, 208
- branch staff establishments, 195
- Brook, R.*, 126
- budget procedure, sales, 164
- cancer and smoking, 124
- ceramic components, sorting, 123
- children of employees, estimation of number, 210
- Clapham, J. C. R.*, 208
- Clarke, G. M.*, 199
- Clifford, P. C.*, 207
- coal industry, 69
- Cochrane, A. L.*, 211
- 'Concise History of Mathematics, A', review, 131
- controlled environments, experimental design, 199
- coronary artery disease, 211
- cost accounting, use of productivity index, 12
- Daniell, P. A.*, 90
- Davies, G. P.*, 47, 209
- Davies, O. L.*, 71, 72
- decisions, statistical, 124
- 'Design and Analysis of Industrial Experiments, The', review, 71
- design of experiments, for controlled environment, 199
- 'Design of Social Research, The', review, 128, 212
- Desmond, D. J.*, 68
- discriminant function, 108
- Doll, R.*, 124
- Dowson, G. H.*, 126
- Dudding, B. P.*, 69
- economic assessment of new industrial methods, 123
- Edey, H. C.*, 134
- 'Elements of Statistics', review, 127
- Emmerson, T.*, 67
- Enrick, N. L.*, 130
- Evans, U. R.*, 209
- Ewan, W. D.*, 210
- experimentation for process improvement, 208
- failure in service, 124
- figures and facts, 69
- foodstuffs, subjective appraisal, 123, 145
- forecasting of sales, 162
- Franklin, N. L.*, 129
- Fry, J.*, 70
- Fryer, H. C.*, 127
- fuel consumption of motor buses, 47
- furnace cycle times, 209
- general practice (medical), 70
- General Register Office, 70
- Glaskin, A.*, 119
- Goobey, G. H. R.*, 126
- 'Handbook of Graphic Presentation', review, 127
- Hardman, A.*, 68
- Harland, S. C.*, 212
- Harper, R.*, 123, 145
- Hartley, H. O.*, 72
- Hepburn, J. H.*, 173
- Hotelling's T^2 , 107
- 'Hulton Readership Survey', review, 184
- industrial data, 69
- input-output tables, 139
- Institute of Incorporated Practitioners in Advertising, 184
- inventory control, 65, 124
- investigation, development, 68
- investment, use of statistics, 126
- James, A. N.*, 123
- Japanese journal 'Statistical Quality Control', 120
- Jessop, Marjorie*, 73
- Jowett, G. H.*, 32, 80, 127
- Keen, Joan*, 68, 69
- Knowles, Elsie A. G.*, 123

APPLIED STATISTICS

- labour turnover, 210
- Lane, K. F.*, 124
- Liddell, F. D. K.*, 90
- life-test data, 68
- linear programming, 22, 209
- Liston, D. J.*, 162
- local statistics, 208
- Logan, W. P. D.*, 70
- Lomax, K. S.*, 124
- Lyle, P.*, 207
- Mahalanobis's D^2 , 111
- makeshift methods, 47
- Maverick, L. A.*, 130
- medical field surveys, 211
- Medical Section of RSS, 207
- medical statistics, teaching, 70
- metallurgy, statistics in, 209
- Monte Carlo methods and communication problems, 208
- Morris, J. N.*, 211
- Moser, C. A.*, 129
- motor industry, 68
- Muir, A.*, 68
- multiple regression, 80
- multivariate analysis, 103
- 'National Income and Social Accounting', review, 134
- national income statistics, 70, 133
- 'National Readership Survey, The', review, 184
- Neale, A. B.*, 132
- non-orthogonal experiments, 210
- non-parametric test, 97
- normality, effect of departures from, 125
- Norris, K. P.*, 123
- operational research, 67, 68, 69
- Padley, R.*, 132, 208
- particle size, 209
- Peacock, A. T.*, 134
- Pearson, E. S.*, 72
- personnel management, 122
- plane sheet, multiple regression, 80
- pre-eclampsia, 97
- presentation of statistics, 70
- price index numbers, 64
- probability in industry, 123
- production control, 9
- 'Productivity, a Critique of Current Usage', review, 130
- productivity in cotton industry, 67
- productivity index, 1
- 'Quality Control through Statistical Methods', review, 130
- quality level for minimum cost, 173
- Ramm, B. T.*, 195
- ratio paper, 16
- Read, D. R.*, 71
- readership of newspapers, etc., 183
- retail trade statistics, 211
- Rivett, B. H. P.*, 69
- Sakura, Itasu*, 64
- sales figures, graphing, 15
- sales forecasting, 162
- sampling, 120
- Saunders, C. T.*, 133
- Schmid, C. F.*, 127
- Sears, G. W.*, 47
- serial variation, 32
- Shankleman, E.*, 183
- significance of difference, non-parametric test for, 97
- , test in time series, 32
- Smith, C. A. B.*, 212
- smoking and lung cancer, 124
- social accounts, 134
- Solt, R.*, 209
- sorting machine for ceramics, 123
- staff establishments, 195
- Stafford, J.*, 211
- 'Statistical Analysis in Chemistry and the Chemical Industry', review, 129
- stock control, 65
- Struik, D. J.*, 131
- subjective appraisal of foodstuffs, 123, 145
- Swan, A. W.*, 15
- television, measurement of audience, 126
- times series, comparison of means, 32
- Tippett, L. H. C.*, 67, 127, 130, 131
- Todd, G. F.*, 70
- Tress, R. C.*, 70
- tuberculosis, 73
- van Rest, E. D.*, 66, 120, 130
- viscose rayon yarn sales, 15
- Warwickshire, tuberculosis statistics, 73
- Williams, N.*, 22, 209
- Wishart, J.*, 103

Printed in Great Britain at
THE KYNOCH PRESS
Birmingham

